asociación española
ingenieros de telecomunicación

Asociación
de Enxeñeiros
de Telecomunicación
de Galicia

PROCEEDINGS

Santiago de Compostela 1 - 4 September 2010

FITCE

49th FITCE CONGRESS

49th FITCE CONGRESS

PROCEEDINGS

Santiago de Compostela 1 - 4 September 2010

Organize:

asociación española
ingenieros de telecomunicación

Asociación
de Enxeñeiros
de Telecomunicación
de Galicia

FITCE

Sponsored by:

XACOBEO 2010

XUNTA DE GALICIA

EDNON

ERICSSON

Gradiant
CENTRO TECNOLÓXICO DE TELECOMUNICACIÓNS DE GALICIA

IBM

telsis

R

Telefónica

Televes

vodafone

Table of contents:

Table of contents:

Andrea Penza
President FITCE Europe

Welcome to the 49th FITCE Congress

Santiago de Compostela 2010

Dear Ladies and Gentlemen,

Distinguished Guests, Congress Participants and Fitce members,

It gives me great pleasure to welcome you to the 49th Fitce Congress in Santiago de Compostela.

The theme of the Congress is "The way of Santiago and European Telecommunications".

This is a very special year for Santiago and it is really important that most of telecommunication topics, especially these ones referred to the future of telecommunication market could be discussed and analyzed in a Spanish area where the perspective of new vision and a new age in telecommunication sector can be concretized despite the still existing economical crisis. When we refer to 'The Way of Santiago" we are taken to think to something new, something which moves on, something good which has to occur and can be able to transform and improve the whole life of each of us. During the existing time we really need to think about a new way to proceed, a new way to think, and a new way to conceive the future of our telecommunication world.

During the Congress we will deal with 'The Broadband Way', as the only way that can lead our society to a prosperous future, 'The new Energy Way' which should face successfully all the forthcoming problems of energy lack and climate crisis, 'The new Economics Way, in order to build a new economy fully based on eco-sustainable criteria.

At the end of the Congress a clearer vision of what the future may bring and how the evolution of Networks and services could really generate a meaningful impact on our future lives, changing ICT industries and transforming Society, will be achieved.

The International Fitce Congress will, as always, enable Fitce members, Congress participants and distinguished guests, exchange data, ideas, views, knowledge and practices on a great variety of technical, business, commercial, financial, legal and social issues as well as to strengthen interactivity, co-operation and friendship.

The papers to be presented have been carefully selected by the International Scientific Committee, and I would like to thank all the members for their professionalism in putting together a diverse, challenging and interesting Congress program.

I would like to thank all the sponsors and patrons for their invaluable contribution to this Congress, without which the Congress could not take place. In particular, I would like to thank the Fitce Spanish and the Fitce Galician Congress Executive Committee, for their professional and enthusiastic volunteer dedicated and tireless work in preparing the Congress over the past 12 months.

I wish all of you an exciting and successful annual Fitce Congress and a wonderful time in Santiago.

Kind Regards

Miguel Merino
Chairman of Local Organizing Committee
Vice President Telecomm Engineers Galician Association (AETG)

Welcome to the 49th FITCE Congress

Santiago de Compostela 2010

Dear Ladies and Gentlemen, most distinguished Guests, Accompanying Partners, all Congress participants, and fellow FITCE members,

I am proud to be able to welcome you to Santiago de Compostela this year.

This 49th FITCE Congress will focus on the present state and future plans for the EU's telecommunications infrastructures under the slogan "The Way of Santiago and European Telecommunications". A slogan which links one of the most ancient means of communication (the Way of Santiago – main link between medieval Spain and Europe) to the most modern ones, the proliferation of telecommunications.

The Way of Saint James is an important vehicle of cultural and religious communication between the people of Europe and a shared symbol among the countries and regions comprising the European Union. Since the Middle Age it has been used by pilgrims on their journey to Santiago de Compostela, the city which reveres the relics of James the Apostle.

The Congress is organized by the Spanish and Galician Associations of Telecommunications Engineers, and is proud to have the backing of the Galician Regional Government in special Xacobeo 2010 organization and the sector's indeed Spanish leading ICT companies. This will facilitate the participation of top European, Spanish and Galician officials and keynote speakers from the telecommunications sector.

We are also privileged to have the backing of the Spanish Universities, especially that of the Vigo School of Telecommunications Engineers (Pontevedra-Galicia) and the support of the organizations comprising Spain's telecommunications engineers and Professional Association of Galician Telecommunications Engineers (COETG).

FITCE 2010 received 80 paper submissions from 15 countries on Europe and US. 38 papers were published and presented as full papers, i.e. completed work (6 pages/20' oral presentation). Furthermore, 15 contributions were accepted and presented as posters. These selection positively contributed to reinforce the overall quality of the Congress and to provide a deeper understanding of the ICT field.

Papers published in each session describe state-of-art research work that is often oriented towards real world applications and highlight the benefits of ICT for industry and services.

The program for this congress required the dedicated effort of many people. Firstly, we must thank the authors, whose research and development efforts are recorded here. Secondly, we thank the members of the technical committee for a valuable help with their expert reviewing of all submitted papers. Thirdly, we thank the invited speakers for their invaluable contribution and for taking the time to syntheses and prepare their talks.

Finally, special thanks to all the members of the AETG and AEIT organizing committee, for the success of this conference.

Three best paper awards are given at the closing session to outstanding papers presented at the conference: an award for the best paper, other for best presenter, plus an award for the best young presenter. The selection is based on the classifications and comments provided by the Technical Committee and also on the oral presentation quality, assessed by session chairs.

We wish you all an exciting congress and an unforgettable stay in Santiago de Compostela.

Keynotes Lectures

Thursday 2
15:00 – 15:40
Room Obradoiro

European R&D shaping the future Internet-enabled world

Luis Rodriguez Roselló
Director a.i. · Converged Networks and Services ·
D G Information Society & Media · European Commission

The new Treaty of Lisbon gives higher impetus to Research & Development, and recognises the key role that RTD plays much beyond the mere boost to industrial competitiveness, as it is now fully enshrined as an essential element of the knowledge-based economy and future wellbeing of Europeans. Furthermore, the new policy coined as the "European Digital Agenda", underlines the importance of Information & Communication Technologies, and particularly of the Internet, for the European economy and society as the "fuel" for innovation and growth.

In this context, the launch by the European Commission of the new Work Programme for Research, including an initiative for a Future Internet Public Private Partnership introduces a new perspective. With a commitment to boost Future Internet infrastructure, services and applications, the new media, service and networking industry and research stakeholders will have a golden opportunity to develop and experiment innovative services in close collaboration with the user sectors so as to lead the way in the mainstream Internet-enabled economy, enabling smarter infrastructures and society at large. This initiative is fully part of the new plans for Research in ICT on the future "pervasive and trusted network and service infrastructures", and complements with a more user driven approach the "traditional" European R&D landscape of longer term research.

Luis Rodríguez-Roselló holds a degree of Telecommunications Engineering (Universidad Politécnica de Madrid). After some years of professional activity as engineer at a private company and as full professor at the Faculty of Telecommunications and Engineering Madrid in Computer Science and Control Systems, he was appointed Director of the R&D Department at the ITE (Institute for Technologies in Education) of the Ministry of Education in Spain and later on Head of the International Department of the CDTI (Centro para el Desarrollo Tecnológico e Industrial).

He joined the European Commission in 1989 as Head of Division in Directorate-General "Information Society and Media" responsible for the R&D Programme DELTA (Developing European Learning through Technological Advance). He also headed the Educational Multimedia Task Force of the European Commission. He was acting Director in 2003 and 2004 of Directorate "Emerging Technologies, Infrastructures & Applications". Domains of responsibility encompassed basic research (Future and Emerging Technologies), Grid Technologies, Research Infrastructures and application areas related to eInclusion and eWork. End 2004 he was appointed Head of the Unit "Networked Media Systems", where he leads European R&D on this topic within the current EC Framework Programme. In November 2009 he was appointed Director a.i. of "Converged Networks & Services".

Keynotes Lectures

Friday 3
15:00 – 15:45
Room Obradoiro

The development of the Information Society, an ongoing task

Daniel Torres Mancera
Director of Spanish Telecommunications and Information Society Observatory. Red.es

Information and Communication Technologies are the very tools that advanced economies use today to secure their long-term leadership in a global world. No surprise then that Information Society policies are a fundamental part of public and private agendas in all developed countries. As new phases of this ongoing process are deployed, parallel, deep, rigorous approaches are needed to be tackled from a global perspective. A large number of international entities work together to define the next wave of Information Society strategies, a comprehensive task to which Spain is a decisive contributor. The Presidency of the European Council in the first half of 2010 has been the perfect framework for the presentation of a number of proposals from Spain in this area. These proposals are the result of a fruitful collaboration with entities such as the OECD, ITU and the European Commission. In Spain, Plan Avanza, the National plan for the development of the Information Society, has achieved remarkable results in its first five years and has set the pace for a second phase beginning next year 2011. Avanza's milestones are to be reinforced in the years to come, as Spain moves forward to overcoming the challenge of turning into a knowledge-based economy and society in the 21st century. As it has successfully been shown by recent projects, international collaboration will prove a key asset in achieving the ambitious goals set so far.

Daniel Torres Mancera, Telecommunications engineer and Executive MBA from Instituto de Empresa and ISEM (University of Navarra), since 2008 he is the director of the Spanish National Observatory for Telecommunications and the Information Society. From 2004 to 2008 he was an advisor to the Secretary of State for Telecommunications and the Information Society, and member of the board of directors of Inteco, S.A. (public company developing accessibility and e-confidence technologies). Prior to that, he was a strategic advisor to CEOs in the international ICT sector. He is a lecturer in marketing, finance and strategic development at business schools and academic forums. He has also worked as international co-operation projects director for the Spanish Human Rights Association. He has worked for fifteen years on projects in Spain, Europe and Latin America in the areas of Information Society (eHealth, eEducation, eGovernment), business development, projects development, eBusiness and ICT and Human Capital. He has started up audio-visual and communications businesses and directed due-diligence processes for investment projects in multinational environments.

Full papers

A5/1 ROBUSTNESS IMPROVEMENT
Randomization of Dummy Bits

Pilar Molina, David Valerdi, Alberto Gómez, Kyriakos Exadaktylos
Vodafone Technology Networks - 2G Product, Vodafone Group, C/ Isabel de Colbrand, Madrid, Spain
maria-pilar.molina@vodafone.com, david.valerdi1@vodafone.com, alberto.gomez@vodafone.com,
kyriakos.exadaktylos@vodafone.com

Keywords: GSM security; A5/1 countermeasures; randomization of dummy bits (RDB); random fill bits (RFB); ciphering enhancements;

Abstract: A hacker presentation given on December 2009 attracted substantial attention from the media on the possibility to break A5/1 encryption. Although hackers have made progress, this is not news as various hacking groups have demonstrated in the past that A5/1 encryption can be broken quickly and efficiently. Vodafone has been developing a plan of innovative A5/1 countermeasures oriented to improve ciphering robustness in the sort term at a minimum cost so that they can be rapidly deployed over current operating networks. One of the most significant mechanisms that have been covered under the scope of this work frame is the one known as "Randomization of Dummy Bits" (a.k.a. Random Fill Bits). This countermeasure has been approved in the GERAN plenary and will be mandatory for handsets from R6 onwards and preferred for networks. In order to assess the benefits derived from this innovative technology, Vodafone has performed several trials in collaboration with different vendors during the last year. As a result of this testing, an enhancement over the basic countermeasure has been developed that will facilitate an earlier commercial deployment. This paper presents the main findings and conclusions obtained from these trials.

1 INTRODUCTION

GSM technology is undoubtedly the most popular and successful commercial cellular mobile communication system in the world. In over twenty years of development, GSM has become the de facto wireless telephone standard, conquering Europe in the first place and extending afterwards to the rest of the world. Nowadays, over 3 billion people in more than 200 countries have access to GSM services (GSMA, 2010) and this number keeps growing rapidly as a result of its success also in emerging markets.

All traffic generated by this huge amount of users needs to be carried over the radio interface in the first place, which implies an obvious drawback for security. The potential threat from eavesdropping this kind of transmissions is relatively high compared to fixed systems that require physical access to capture the communication. Radio link can be easily intercepted and therefore it is very important to have implemented reliable encryption mechanisms to ensure the privacy of GSM users.

Encryption in the GSM system is based on a stream cipher known as the A5 algorithm. This algorithm scrambles the user's voice and data traffic between the handset and the base station to provide privacy, and it is implemented in both the handset and the base station subsystem (BSS). Multiple versions of the A5 algorithm have been defined, implementing various levels of encryption (GSM security, 2010).

This paper is focused on the A5/1 algorithm, which is the first and most extended ciphering standard for GSM system, currently under challenge. The following pages present an analysis conducted by Vodafone on the benefits of one of the most significant countermeasures that have been proposed in the industry to enhance the performance of the A5 family of ciphering algorithms, known as "Randomization of Dummy Bits" or "Random Fill Bits". This analysis is based on the results obtained from multiple trials that have been performed during the last year in collaboration with several main network vendors.

The paper is structured as follows. First, an overall review on the A5/1 algorithm is given in the following Section II. Then, the concept underneath

Randomization of Dummy Bits feature is introduced in Section III and through the next section the performance analysis process is exposed. Finally, main conclusions are presented in the last section V.

2 AN OVERVIEW ON A5/1

Proper ciphering mechanisms are critical in radio communications to protect data from interception. For that purpose the GSM standard specifies the A5 algorithm, used for encrypting the speech, data and signalling information over the radio link.

In brief, GSM system uses symmetric cryptography, which means that data are encrypted using an algorithm, that uses a ciphering key (Kc) as input. This same Kc is also needed by the decryption algorithm to decrypt the data. The idea is that the Kc should only be known by the terminal and the network in order to make data meaningless to anyone potentially intercepting it.

At a given time, the network can command the terminal to start ciphering the data (once authentication process through non standardised A3 algorithm is completed) using the Kc generated before. The network can pick from a number of algorithms to use, as long as the phone supports the one chosen. Handset capability is indicated to the network previously by way of a specific classmark message. The algorithm is also fed with the value COUNT, which is based on the TDMA frame number. The whole scheme is shown in Figure 1 below (Quirke, 2004).

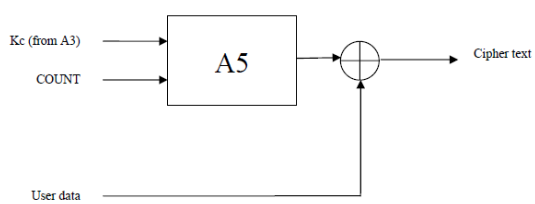


Figure 1: Encrypting data using a ciphering key.

GSM recommendations specify that the network can choose from up to 7 different ciphering algorithms (or no ciphering at all), taking into account that it must be an algorithm supported by the mobile station (ETSI, 2003). Currently there are only 4 algorithms implemented in practice, known respectively as A5/0, A5/1, A5/2 and A5/3. The reason for the diverse implementations is due to export restrictions of encryption technologies.

A5/1 is the original algorithm defined by the GSM standard, a stream cipher used for encrypting over the air transmissions. Although it was developed in 1987 it is still widely used in Europe and America. A5/2 was a deliberate weakening of the algorithm for certain export regions and it is commonly used in Asia. Countries under UN Sanctions and certain third world countries use the A5/0, which comes with no encryption. A5/3 was added in 2002 and is based on the open Kasumi algorithm defined by 3GPP. This algorithm is the stronger option, but it is still on an initial deployment phase (SANS, 2010).

A5/1 and A5/2 were initially kept secret, but the details of their implementation were partially disclosed in 1994 and finally its exact design was reverse engineered by Briceno from a GSM phone in 1999 (Briceno, 1999). Since then, serious weaknesses have been detected in both algorithms: it is now possible to break A5/2 in real-time with a ciphertext-only attack, and on 28 December 2009 German computer engineer Karsten Nohl announced that he had cracked the A5/1 cipher as well (Nohl, 2009). All these recent threats show that the time has come to review and strengthen current ciphering mechanisms on GSM networks, in order to keep user's privacy safe. Network operators need short term solutions that can be easily applicable and robust enough to resist these potential attacks until newer and stronger algorithms (like A5/3) become widely deployed.

3 IMPROVING A5/1: RANDOM DUMMY BITS

In order to face all A5/1 security vulnerabilities, Vodafone has developed a plan of innovative A5/1 countermeasures that aim to make eavesdropping on an A5/1 network significantly harder. These enhancements are oriented to improve ciphering robustness in the sort term at the minimum cost so that they can be rapidly deployed over current operating networks. Vodafone is the owner of an international patent on some of these features.

One of the most significant countermeasures that have been developed under the scope of this work frame is the one known as "Randomization of Dummy Bits" (a.k.a. Random Fill Bits), which basically consists in reducing the predictable information sent in the signaling messaging, both in the uplink and the downlink (Gómez, 2008).

The functionality is based on a simple idea. Fill dummy bits are added to GSM messages when they are sent over the air, if they are smaller than the fixed message size required, as shown in Figure 2 below. The countermeasure RDB reduces the amount of known information sent over the air interface by changing the fixed value previously specified by 3GPP for these fill bits (0x2B) and using a random value instead (3GPP, 2008). Thus, RDB makes it harder to crack the A5/1 encryption algorithm and protects user's communications on the radio interface, helping to reduce the risk for eavesdropping and fraud. In principle, the usage this functionality on the uplink and downlink is independent of each other. BSS controls the usage on the downlink and the MS on the uplink.

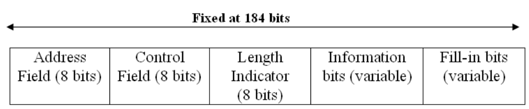


Figure 2: LAPDm frame format.

The major benefit expected from RDB is a significant increase in the system attack complexity by reducing the possibility of a successful probability assessment. As a consequence of this enhancement, the time until the A5/1 algorithm is actually cracked can be significantly delayed; a fact that is especially relevant considering the large footprint of legacy terminals and transceivers that lack A5/3 capability under deployment at the moment.

This countermeasure has been approved in the GERAN plenary and will be mandatory for handsets from R6 onwards and preferred for networks.

4 PERFORMANCE TESTING

In order to evaluate what kind of benefits will be obtained by adopting this innovative technology, Vodafone has performed several trials in collaboration with major network vendors during the last months, both in lab and real network environment. An exhaustive set of test cases was defined for the purposes of this analysis, considering a wide range of services (voice, SMS, GPRS, USSD, etc.) over different network equipment, and studying also terminal compatibility and side effects.

In the first quarter of 2008, a first testing in lab environment was performed in collaboration with a major network vendor (vendor A). These tests covered all kind of functionality verification cases that aimed to deeply assess the overall performance of the feature. No potential impact over network performance was detected and feature worked as expected.

Next logical step in the development plan required feature activation and assessment under real network conditions, so Vodafone prepared and performed a live trial in Navarra during the second quarter of 2009, this time in collaboration with another main network vendor (vendor B). During this testing an issue affecting certain handset platform was experienced after deployment, and it was discovered that some specific terminals presented interoperability problems when activating RDB feature. Affected handset supplier acknowledged the issue and agreed to collaborate and perform new interoperability tests (IOT) with both vendors A and B.

As a result of this tests it was uncovered that some terminals are not totally compliant with the 3GPP standard and do not ignore the fill bits introduced in the LAPDm frames. These mobiles require the pattern 0x2B to be present from the fourth octet on for the corresponding LAPDm frames. The problem has been only observed for SDCCH messages (SAPI=0) when the length of the useful data is equal to zero. Detailed LAPDm frame format is shown in Figure 3 below.

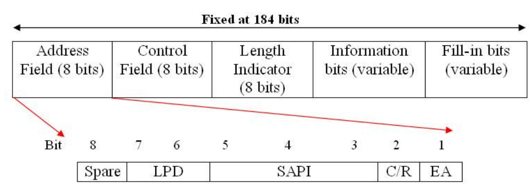


Figure 3: LAPDm frame format – detail of the address field.

In order to solve the problem, an additional requirement has been added to the initial RDB specification, slightly modifying the feature behaviour. For each SDCCH message the first octet of the fill bits (from octet 4 onwards in the LAPDm frame) is set to 0x2B.

These workaround was ad-hoc implemented first, tested and proved effective by collaborating vendors A and B in lab environment during the third quarter of 2009. A commercial implementation of the whole RDB feature (workaround included) was tested by vendor B during the first quarter of 2010 in lab environment in Spain for more than 120 different phones and datacards, considering multiple chipset suppliers as well. No further issues were experienced and all tests cases were passed successfully.

A new trial over real commercial network is planned to be performed in the third quarter of 2010, intended to confirm commercial readiness and feasibility of RDB feature and discard any further interoperability issues when deployment over real networks becomes a reality.

5 CONCLUSIONS

This paper presents an innovative enhancement on A5/1 ciphering algorithm known as Randomization of Dummy Bits (a.k.a. Random Fill Bits) that has been developed and proved effective in practice as a result of the collaboration between Vodafone and several major industry vendors. Vodafone played also a major role in the discussions within the industry that ended up with the inclusion of this feature in the appropriate GERAN specification (TS 44.006).

Results obtained from the multiple tests show the promising performance and applicability of this technology. Furthermore, an improvement over the basic countermeasure has been developed and incorporated in the commercial solution, fact that will undoubtedly facilitate an earlier commercial deployment.

REFERENCES

GSM associaton, 2010. http://www.gsmworld.com

GSM-security, 2010. http://www.gsm-security.net

Quirke, J., 2004. *Security in the GSM system.* AusMobile. http://www.ausmobile.com.

ETSI, 2003. *Recommendation GSM 02.09*

SANS Institute, 2010. *The GSM Standard (An overview of its security)*. http://www.sans.org/reading_room/

Briceno, M. et al., 1999. *A pedagogical implementation of the GSM A5/1 and A5/2 "voice privacy" encryption algorithms*. http://cryptome.org/gsm-a512.htm

Nohl, K. , Krißler, S., 2009. *Subverting the security base of GSM,* https://har2009.org/program/attachments/

Gómez, A., Arranz, M., 2008. *Method for improving the security in GSM networks*. European Patent Application No. 08380067.2

3GPP, 2008. *Technical Specification 04.06 V8.4.0*

Assessment of the exposure to electromagnetic fields produced by high power open industrial microwave systems

Eva M. Muñoz#1, A. Tatiana Zona#2, Juan V. Balbastre#3, Elías de los Reyes#4

Instituto de aplicaciones de las tecnologías de la información y las comunicaciones avanzadas (ITACA), Universidad Politécnica de Valencia, Camino de Vera s/n, 46022 Valencia, España

[1]evmuna@itaca.upv.es

[2]tatianaz@itaca.upv.es

[3]jbalbast@itaca.upv.es

[4]ereyes@dcom.upv.es

***Abstract*— Industrial microwave systems are of increasingly interest in many areas, among them the disinfestation and drying of wood timber and fine arts items. Although in this field microwaves are usually cleaner, more effective and less toxic than the conventional chemical treatments, the exposure to electromagnetic fields must be properly assessed. In this work, safety areas around the antenna applicator of a high power microwave wood disinfestation system have been established. These areas have been defined taking into account the exposure limits for electromagnetic fields, so that the person operating the machine is protected. The research was carried out within a shielded semi-anechoic chamber using a circularly polarized antenna prototype, a shielded suit, radiation measurement equipment and some wild pine wood samples. The radiated electric field strength was measured around the prototype and three radios defining the safety areas were calculated. The results show that the operation of such kind of microwave machinery fulfills the regulatory limits on the exposure to electromagnetic fields of both people and workers provided that appropriate safety areas are established and respected.**

I. Introduction

The use of microwave energy for heating was first reported by in 1945 by Percy Spencer. One of the earlier applications found in the scientific literature was to use microwaves as an alternative to chemical pesticides in agricultural products [1-4]. Microwaves are effective against insect pests due to the extremely different moisture content of grain (~10%) and insects (>80%), hence producing the temperature of the insect to rise over lethal values. In the 1990's Andreuccetti et al. [5-7] proved the effectiveness of microwave to disinfest wood attacked by woodworms.

In [8] is presented a linearly polarized microwave wood disinfestation system prototype. The prototype complied with international, European and Spanish recommendations and regulations limiting the exposure to electromagnetic fields if some zoning restrictions are met. The same prototype was used to establish an initial procedure to exterminate woodworms in panel painting [9, 10]. ~~In a~~A circularly polarized applicator wich improves the system performance was ~~resented~~designed. This prototype is property of DESINSECCION Y SECADO EN 3D, S.L. with a trade mark name DRYPARASITE and property of the licence.

This paper is focused on presenting the work carried out in order to define safety areas for both general people and workers during microwave wood treatment using the aforementioned prototype, so that the compliance with exposure limits to electromagnetic fields may be guaranteed. To achieve this, the field strength around the microwave applicator was measured on planes parallel and orthogonal to the antenna aperture. Then the attenuation of a shielded suit was measured to determine the safety area where only authorized people properly dressed can access and finally the area where nobody can stay even wearing special clothes was established.

II. Regulation

Exposure to a radiofrequency (RF) field of sufficient intensity can cause heating, and the limits and guidelines for human exposure are set to prevent a temperature rise in the body.

European Directive 2004/40/EC on occupational exposure to electromagnetic fields (EMF) [11], based on the guidelines of the International Commission on Non-Ionizing Radiation Protection (ICNIRP), was published in 2004. The directive was to be implemented in the Member States of the European Union (EU) by 2008. However, because of some unexpected problems, the deadline was postponed until 2012. European exposure standards and drafts harmonised with Directive 2004/40/EC are under development to help both manufacturers and employers to comply with the basic safety requirement and to provide the protection of the health and safety of the worker and any other person in the proximity of an electromagnetic field.

The Directive defines Action Values, which are the limits of exposure based on variables that can be directly measured; the electric field strength E, the magnetic field strength H, the magnetic flux density B and the power density S. For working environments that are also accessible to members of the public, the assessment of exposure levels is already available in accordance with the EU recommendation for limiting the exposure of the general public [12]. In that case, the values that limit the exposure are defined as Reference Levels.

European Standard EN 50499:2008 [13] provides a general procedure in order to assess workers' exposure to electric, magnetic and electromagnetic fields in a work place to demonstrate compliance with exposure limit values and action values as stated in the Directive 2004/40/EC. The exposure to be assessed is at the work place(s) where a worker is permitted to be present. Once the equipment to be used in the work place has been identified and its characteristics and normal conditions of use established, an appropriate method may be used to demonstrate compliance with Directive 2004/40/EC. Some existing standards [14], [15] and drafts can be useful to establish that method in order to obtain the assessment of the risk results. A zoning concept (definition of different areas at the work place, in terms of the exposure level and limits) is also explained in EN 50499.

III. EXPERIMENTAL

The establishment of safety areas around the microwave disinfestation system is based on the determination of boundaries where the exposure limits might be exceeded. In that way, the generated electromagnetic field when the system is working has been measured and evaluated.

The system prototype has two microwave sources independently controlled, so it can be operated on two modes: in Mode 1 only one source is working, in Mode 2 both sources are working. The sources are connected to the applicator through high power coaxial cables. The applicator is made up of a section of square waveguide, suitable for operating at 2.45 GHz and it radiates circular polarized fields.

The measurement of the electromagnetic fields was carried out using a radiation detector along with a 100 kHz to 3 GHz electric field probe. The measurement probe must be placed in the far field region, which means at least 37.7 cm away from the radiating aperture.

The tests were carried out in three stages. In the first one, the applicator was put on the center of a circular turntable and the radiation was directed towards the floor. In order to obtain the radiation pattern around the applicator, the electric field was measured on several planes parallel to the chamber floor, at four different heights: 0 m (on the surface), 1 m, 1.5 m and 1.75 m. The field probe was placed 2.5 m away from the center of the turntable and the electric field was measured each 30º while a wood sample was irradiated during 20 s first in Mode 1 and next in Mode 2 of operation.

To continue in the second stage, the applicator was placed on the turntable surface without a wood sample. In order to establish the radiation pattern, the electric field was measured 2 m away from the radiating aperture each 45º while the antenna was radiating to the air. The measurements were done on three different planes parallel to the floor: one of them corresponding to the line of sight of the applicator and the other two were placed ±50 cm from the first one.

The last step (stage 3) aimed to assess the attenuation of a shielded suit, so the methodology was identical to the second one described in the previous paragraph, but now the field probe was attached to the hip, chest and head of a manikin dressed with a shielded suit. In this way, the measures have been carried out at some representative heights corresponding to critical tissues in a human body. Apart from the distance of 2 m from the applicator aperture, another distance of 1.5 m was considered to do the measures at hip level, in order to estimate the relative error of the calculation.

Once the electric field was measured at a distance D, it was calculated where the exposure limit is exceeded:

$$R = \sqrt{\frac{(D(m))^2 \times (Emax(V/m))^2}{(Limit(V/m))^2}} \qquad [1]$$

Based on the results of the measured electric field, three boundaries can be defined to delimit the areas where a person may stay when the system is working: R1 is the distance where the measured electric field complies with the reference levels for general public (61 V/m) [11], R2 represents the distance where the action value for occupational exposure (137 V/m) [12] is achieved and R3 corresponds to the distance where a worker wearing the shielded suit would be exposed to the corresponding action value, taking into account that the measured value is the electric field attenuated by the suit and the limit is again the action value for occupational exposure (137 V/m) [12].

IV. RESULTS

The results were processed taking into account that occupational exposure assessment should always consider the worst-case situation.

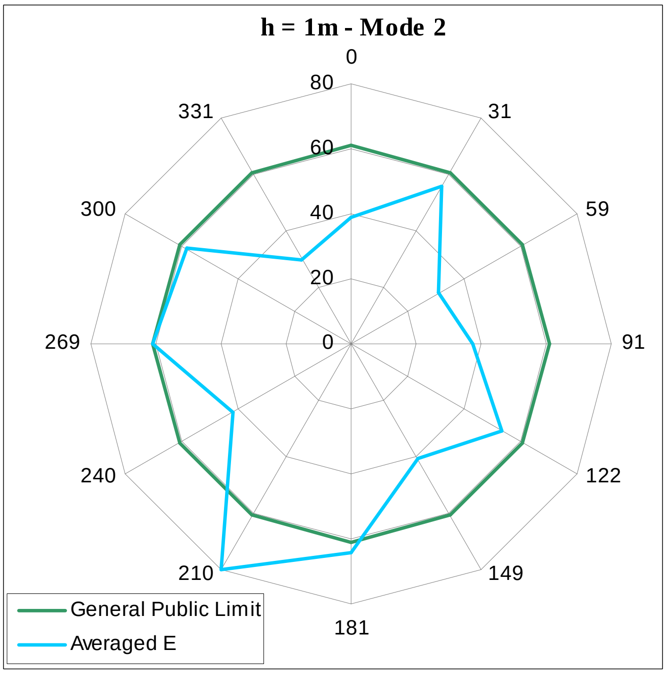


Figure 1 – Electric field in the surface h=1m for operation Mode 2

The boundaries defining safety areas around the applicator in planes parallel to the surface of radiation (stage 1) were calculated. The most restrictive situation was found for a treatment carried out with sources in Mode 2, at a height of 1m and angle of 210º, as it is shown in Figure 1. This is, besides, the only result in which the limit for public exposure was exceeded; the measurements done at the rest of planes where always below the reference level.

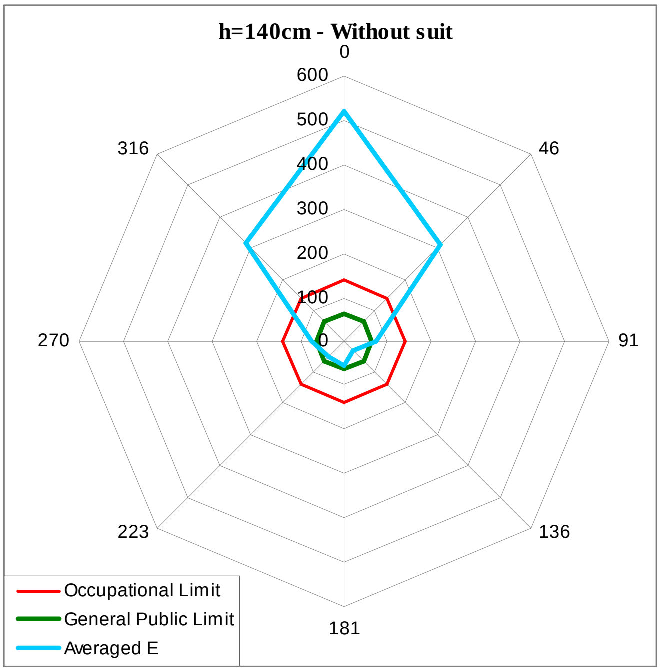


Figure 2 – Radiated electric field measurements without suit at h=140cm

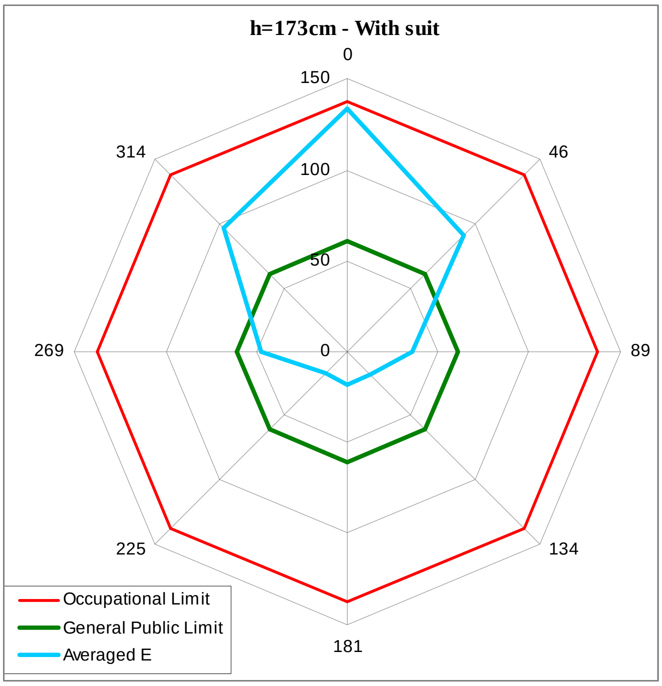


Figure 3 – Radiated electric field measurements with suit at h=173cm

Then, all measurements achieved at Stage 2 (without suit) and Stage 3 (with suit) were examined and compared to those in Stage 1, and the higher values of electric field were always found when the applicator radiated directly to the air. Accordingly, this is the worst-case situation. Figure 2 and 3 show the results in which the average of the measured electric field in such case (with and without shielded suit covering the field probe) are most significant. Measured data are called Averaged E and they are compared with the general public and workers (occupational) limits.

Finally, to establish the distances which define the different areas where the operator or general public may stay, it is necessary to identify the most significant ranges of angles from the analysis of the graphics in Figure 2 and 3. R1, R2 and R3 are obtained and detailed in tables 1, 2 and 3 for the worst–case situation of operation (Mode 2).

TABLE I
DEFINITION OF DISTANCE R1

Angles (º)	E max (V/m)	R1 (m)
[285-75]	520.13	17.05
[75-90]&[270-285]	100.00	3.28
[90-270]	55.37	1.82

TABLE II
DEFINITION OF DISTANCE R2

Angles (º)	E max (V/m)	R2 (m)
[315-45]	520.13	7.59
[45-70]&[290-315]	317.33	4.63

[70-90]&[270-290]	137.00	2.00
[90-270]	55.37	0.81

TABLE III
DEFINITION OF DISTANCE R3

Angles (º)	E max (V/m)	R3 (m)
[315-45]	137	2.00
[45-70]&[290-315]	89.61	1.28
[70-90]&[270-290]	60.00	0.88
[90-270]	47.02	0.67

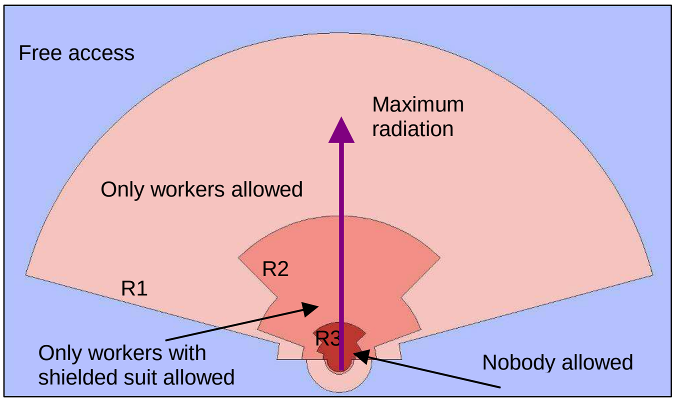


Figure 4. Accesibility areas for operation mode 2

These boundaries are plotted in Figure 4 to show the safety areas around the applicator when the radiated aperture plane is placed perpendicular to the floor and the applicator is radiating to the free space.

It must be taken into consideration that the distances R1, R2 and R3 shown in Table I, II and III depend on the type and thickness of the wood piece to be treated, and the treatment applied. As they have been estimated in the worst–case situation, the boundaries found in this work are considered valid for any type of treatment.

V. CONCLUSIONS

A method for obtaining the safety areas around a system for microwave disinfestation of wood has been proposed and described, based on the measurements carried out in several locations at different planes from the applicator. Through this procedure, the points where the exposure to electromagnetic fields is critical for both general public and workers can be assessed and restricted areas may be delimited in order to protect the user of the device or the general public who might stay in the proximity of the working zone.

The established accessibility areas are valid only for the wood disinfestations system under test, therefore, another radiated power or polarization might generate a different outcome.

The results have demonstrate that it is feasible the use of an open horn-like applicator radiating at 2.45 GHz. The feasibility is considered from the point of view of compliance with national and European regulation limiting human exposure to electromagnetic fields, particularized to an occupational environment.

Additionally, for close (handheld) applications the worker must wear special shielded suits. Although it seems that those suits do not offer enough protection at very short distances (below 0.5m, approximately), these results have been obtained using a suit with a shielding effectiveness figures below 13 dB, and it is important to emphasize that nowadays, in the market, there are readily available tissues offering shielding effectiveness bigger than 30 dB at 2.45 GHz. Suits made of those tissues will allow the worker to manually operate the applicator.

ACKNOWLEDGMENT

The authors would thank the Spanish company Desinsección y secado 3D S.L., particularly Antonio Martinez, who has supported the work presented in this paper. For further information contact Info@dryparasite.com

REFERENCES

[1] A. Díaz-Morcillo, P. J. Plaza-González, J. V. Balbastre, J.M. Osca and D. Sánchez-Hernández, “Commercial rice disinfestation unit using microwave energy”. *Proceedings of 9th International conference on microwave and high frequency heating, 2003*.

[2] S. L. Halverson, W.E. Burkholder, T.S. Bigelow, E.V. Norsheim and M.E. Misenheimer, “High-power microwave radiation as an alternative insect control method for store product”. *Journal of economic entomology*, 89(5): pp 1638-1648, 1996.

[3] S.O. Nelson, “Review and assessment of radiofrequency and microwave energy for store-grain insect control”. *Transactions of the ASEE*, 39(4): pp. 1475-1484, 1996.

[4] H.H. Webber, R.P. Wagner and A.G. Pearson, “High frequency electric fields as lethal agents for insects”. *Journal of Economic Entomology*, 39: pp 487-498, 1946.

[5] D. Andreuccetti, M. Bini, A. Ignesti, A. Gambetta and R. Olmi, “Microwave destruction of woodworms”. *Journal of Microwave Power and Electromagnetic Energy*, 29(3): pp 153-160, 1994.

[6] D. Andreuccetti, M. Bini, A. Ignesti, A. Gambetta and R. Olmi, “Feasibility of microwave disinfestations of wood”. *Proceedings of 26th annual meeting IRG/WP*, Doc no. IRG/WP/95-40051, June 1995.

[7] D. Andreuccetti, M. Bini, A. Ignesti, R. Olmi, S. Priori, A. Gambetta and R Vinni, “A microwave device for woodworms disinfestation of artistic painted boards”. *Proceedings of the international conference on microwave and high frequency heating*. 1995.

[8] P.J. Plaza, A.T. Zona, R Sanchis, J.V. Balbastre, A. Martínez, E.M. Muñoz, J. Gordillo and E. de los Reyes, “Microwave disinfestation of bulk timber”. *Journal of Microwave Power and Electromagnetic Energy*, Vol. 41 no. 3 pp 22-37, 2007.

[9] A.T. Zona, J.V. Balbastre, L. Nuño, E. de los Reyes, V. Vivancos and E. Perez-Marín, “Initial procedure to exterminate woodworms in panel paintings by microwave-power application”. *Proceedings of IMPI’s 40th annual symposium*, pp 267-271, August 2006.

[10] V. Vivancos, E. Perez-Marín, L. Nuño, J.V. Balbastre, A.T. Zona, “Microwave treatment dor woodworm disinfestion in large-format Works of art”. *Proceedings of the international conference on heritage, weathering and conservation*,Vol. II pp. 707-712, June 2006.

[11] 1999/519/EC, Council Recommendation of 12 July 1999 on the limitation of exposure of the general public to electromagnetic fields (0 Hz to 300 GHz).

[12] Directive 2004/40/EC of the European Parliament and of the Council of 29 April 2004 on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (electromagnetic fields) (18th individual Directive within the meaning of Article 16(1) of Directive 89/391/EEC).

[13] Procedure for the assessment of the exposure of workers to electromagnetic fields. EN 50499:2008.

[14] Basic standard on measurement and calculation procedures for human exposure to electric, magnetic and electromagnetic fields (0 Hz - 300 GHz). EN 50413:2008.
[15] Assessment of electronic and electrical equipment related to human exposure restrictions for electromagnetic fields (0 Hz - 300 GHz). EN 62311:2008.

SERVICE ASSURANCE IN FIXED BROADBAND ACCESS NETWORKS

Optimization in network restoration problem

Martín J. Alarcón, Diego Andina, Juan B. Grau, Joaquín Torres

E.T.S.I.Telecomunicación & E.T.S.I.Agrónomos, Universidad Politécnica de Madrid (UPM), Madrid, Spain

martin@gc.ssr.upm.es, d.andina@upm.es, j.grau@upm.es, joaquin@gc.ssr.upm.es

Keywords: Service Assurance, broadband access networks, ELECTRE, outranking relationships, multicriteria, alternatives.

Abstract: In the access network there is no alternative routing: local loop of a residential customer is connected to a single port of single equipment. If a fault appears there, a technical operator has to move to repair the fault. Failures in the access equipment can occur at any point in the network, and obviously the amount of available technical operators in a company is limited. Therefore an algorithm to optimize travel routes for each operator would save a lot of money to the company. The present paper describes an algorithm used to solve this optimization problem based on discrete multicriteria method ELECTRE I.

1 INTRODUCTION

Nowadays, large telecommunication companies doesn't earn money with traditional voice traffic. In recent years the major source of income is broadband services, thanks to the mass use of high speed internet access. Following the internet service is also growing increasingly the use of IPTV and VoIP services.

Large companies invest more and more in their own network infrastructure. They buy new equipments, more sophisticated and faster than current ones, and they install network redundancy to have alternative routing. All these actions improve network capacity and aim to obtain a fault tolerant network. Companies want to ensure a high quality of service for final customers in order to maximize customer satisfaction.

However, the broadband services to residential customers, through a fixed line access, are characterized by a single access point to the network. In the access network there is no alternative routing: local loop of a residential customer is connected to a single port of single equipment (DSLAM). And in the new passive optical access networks (GPON), enabling broadband services much higher bit rate, same thing happens: the single-hair fiber that reaches customer's house is connected to a single port of optical access equipment. In both cases if the port has a fault then the client service is interrupted. Since the locations at which access equipments are usually installed are not attended by any staff, with the appearance of one of these faults is necessary for a technical operator moves to repair the fault (Hayashi, 1996). The most common repair is to connect the access cable, from current port to another port of access equipment (destination port must be free and operational). The port is chosen by the technical operator with the help of a legacy inventory management system.

Technical operator trips, to do every repair, generate a cost to the company. Failures in the access equipment can occur at any point in the network, and obviously the amount of available technical operators in a company is limited. Therefore an algorithm to optimize travel routes for each operator would save a lot of money to the company. The present paper describes an algorithm used to solve this optimization problem based on discrete multicriteria method ELECTRE I.

ELECTRE I algorithm is based on outranking relationships between decision alternatives. It is necessary to define separate known criteria and to compare the performances of alternatives. Decision maker has to state preference thresholds. The primary purpose of these thresholds is to recognize that the performance of one alternative is strictly preferred to that of another only if there is a sizable difference in their scores.

Without any doubt, a good definition of the functions, that are used to calculate the score of each alternative, is the key to successfully resolve the problem. It is desirable to get repaired all the faults existing in the same location, with a single operator trip. But at the same time, not all failures have the same priority, depends on whether the customer has made a complaint or not for loss of service, and elapsed time since it was made. Also other factors come into play, as the existence of VIP customers, the existence of monetary compensation to customers due to service interruption, etc. The present paper describes all these factors, and explains how these items contribute to the decision problem that the algorithm ELECTRE I intends to solve, in a real case.

2 ACCESS NETWORK

This paper focus on broadband access networks (Cameron, 2004) and the equipments that are installed there.

As we have said before, we can have copper line or optical fiber based networks. Here is a tipical connectivity diagram of an access network made with DSLAMs.

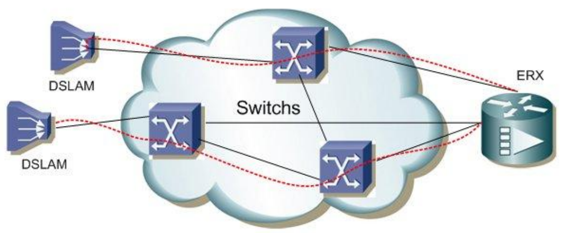


Figure 1: DSLAMs access network.

A Digital Subscriber Line Access Multiplexer (DSLAM) is a network device, located near the customer's location that connects multiple customer DSLs to a high-speed Internet backbone line (through an Edge Router Switch) using multiplexing techniques. It allows local cooper loops to make faster connections to the Internet. By locating DSLAMs at locations remote to the telephone company central office (CO), telephone companies are now providing DSL service to consumers who previously did not live close enough for the technology to work. The ability to provide high speed Internet service from a central location is attractive but the issues of equipment and operational costs exist.

From the physical network infrastructure, a logical network is constructed defining paths of different bit rate. There will be other's paths clients, so that a sufficient "capillarity" is obtained to provide services to all the users (final customers), who ask for it. In the top layer of this hierarchy we have the final services that can be bought by a customer (ITU-T, 2000).

The new gigabit passive optical access networks (GPON) follows the same connectivity scheme, besides enabling much higher bit rate services. In GPON the role of DSLAM is played by Optical Line Termination (OLT) equipment (Yong-Kyung, 2003).

These network architectures allow users to get specific performance guarantees which are defined in a Service Level Agreement (SLA) document (M. D'Arienzo, 2004). SLA represents a formal high level definition (user view) of characteristics for a communication service. Our aim, in the present paper, is to help telecommunication companies to meet SLAs agreed with its customers, by decreasing the customer services unavailability time.

3 INPUT DATA

The operation of the algorithm, presented in this article, needs access to the full database of all customer services. It also needs the list of alarmed services (that is, the customer services that are unavailable because of a network failure).

From each customer service, the algorithm will take into account only the following attributes or characteristics:

- Customer category
- Quality of service agreed
- Bit rate
- Price
- Critical services
- Type of service cut off (VoIP, internet, TV)
- Existence of customer's complaints

These attributes are described in more detail in the following sections.

3.1 Customer category

We will suppose that each customer will have assigned a category that will indicate the volume of business that it has with the operator. The customers with greater volume of business will be more important for the company, since if some of them are displeased with the service and decide to cancel its contract this would suppose an appreciable income fall off. We will suppose that four values of category will exist, from value 0 (residential customers with few privileges) to value 3 (big companies with huge volume of business). There will be a measure of the importance of each category, from the company business point of view. This is shown in the following table.

Table 1: Customer categories.

Category	Description	Importance
0	Residential customer	1
1	VIP client	2
2	Small company	5
3	Big company	20

3.2 Quality of service

Each circuit will have a quality of service contracted by the customer. If the company breaches the quality agreed, it will suffer money penalties.

We will suppose that a customer can subscribe a quality level equal to 1, 2 or 3. If a customer wants a better quality level he will have to pay more for the circuit.

Quality will decrease with the time a service is unavailable.

Therefore, the penalty the company suffers depends on the time each service is cut (usually there will be predefined thresholds of time for this) and each circuit quality level.

The penalty also depends on each service price.

3.3 Bit rate

The bit rate determines the transmission capacity, that is to say, the amount of bits per second able to communicate. Typical bit rates are: 1 Mbit/s, 2 Mbit/s, 10 Mbit/s.

As much as the bit rate is, much is the service price, and therefore, penalties in case of unavailability problems, will be greater for the company.

3.4 Price

The customer services have a price (P) that is monthly considered.

We will assume that the monthly price depends only on two variables:

- Bit rate (B), measured in megabit per second (Mbit/s).
- Quality of service (QoS), whose value is 1, 2 or 3.

Considering these factors, we will suppose that the monthly price of a service follows the formula:

P = (1+B/3) + QoS*B/10

We will use the price in the calculation of penalties if quality of service is not accomplished.

3.5 Critical service

Regardless of client category, certain services will be considered critical because they are essential for some purpose. If the rest of the properties are the same, a service with a critic mark would have to be recovered quicker than a service whose client category is “3” (maximum).

We will suppose that any service can be labelled with a "critical level" value among the following ones:

- level 2: it will be used for official sites circuits, like police station, hospitals, ministry, etc.
- level 1: used for circuits dedicated to the own telecommunication management network.
- level 0: for the rest.

3.6 Service type

In this section we deal with the different type of services that can exists.

We have the next types:

- Basic internet access. It is the broadband service that almost all customers have. It provides IP internet access to customer premises equipments (CPE).
- Voice over IP service (VoIP). It provides a telephone service as if it were an old analog line.
- Television over IP (IPTV). This service demands a high bit rate because of the big amount of data it needs.

Among the three, the most important is VoIP service, because de customer wants it to be 100% available, as old analog one was.

IPTV service is also important for the customers because it tends to be expensive in the bill that the customer pays.

3.7 Customer’s complaints

When a customer detects that one of his telecommunication services doesn't work, he usually makes a complaint to the operator company.

On the contrary, there may be customers affected by a network incidence, who have not realized that, because they are not using the service in that moment.

Obviously, it is more urgent to restore the service of the customers who have made a complaint, with respect to those that have not made it.

It is also important the time that takes from the beginning of the complaint, until the problem is solved, since it influences the sensation of efficiency that the client perceives, with respect to the capacity of the company to solve his problem.

For every circuit affected by a network failure it is necessary to know if it has generated any customer complaint, and if so, it is also needed to know the date and hour in which the customer fulfilled the complaint.

4 MULTICRITERIA METHOD

From multicriteria decision-making methods, ELECTRE I (Roy, 1991) has been chosen for our problem, since it allows to find the best alternative without concerning the relative arrangement of the rest of the alternatives.

The ELECTRE algorithms belong to outranking methods family (Romero, 1993). The ELECTRE approach starts from the intuitively premise that a Decision-Maker (DM) can only make approximate comparisons of the performances of alternatives. ELECTRE allows performances which are not numerically equal to be considered equal. Outranking does not have an axiomatic basis, but rather is based on parameters and a decision algorithm. It is still necessary for the Decision-Maker to provide the analyst with scores for the alternatives against the criteria, but the preference system is “designed” via the approach. Thresholds are set which reflect the DM’s wishes to compare these performances in an imprecise manner.

So, first of all, it is necessary to define the aggregated criteria that are going to be used in ELECTRE I method (Antón, 2004). That is a mission for Decision-Maker staff. The criteria adopted here are the following ones:

- Material Cost to execute the repair. It includes cost of workers’ trips, and material cost of the installation of new equipments, wirings, welds/joints, etc.
- Human Cost to execute the repair. It depends on the amount of involved workers, their salaries and the time that they dedicate to the task.
- Direct Cost of the penalties for not complying with the quality of service subscribed by the affected clients. As commented before, if the time in which a customer has no service overcome a threshold, the operator company has to pay a money penalty.
- Potential income. The customers dissatisfied with the service can decide to finish their contract and to change the company. The first operator company won't get more income from this customer.
- Negative publicity, due to malfunction of the services of a big hospital or of an airport, etc. These "disasters" bring about diminution of the incorporation of new customers to the company, who will go to other companies. This criterion is measured in units of time (minutes). It is the accumulated fail time of all the services whose cut has a great repercussion in media.

Time inverted in the repair has not been considered as an isolated criterion, since it is included in several of the already enumerated criteria (the time always becomes a cost).

The weights assigned to each criterion (again decided by the Decision-Maker) are shown in the next table.

Table 2: Weights assignment.

Criteria	Weights	'more is better'
Material Cost	0,15	-1
Human Cost	0,20	-1
Penalties	0,10	-1
Potential income	0,25	1
Negative publicity	0,30	-1

5 ALTERNATIVES

The next step in ELECTRE I method is to define the decision alternatives.

Each alternative must be a complete solution of arrangement, that is to say, it must provide an ordering rule that allows to be applied to all the services that are affected by an network incidence. To apply an alternative must produce an ordered list of services. The maintenance workers of the company will follow this list to restore each service, in sequence.

If we have N affected services, the possible different ways to order them are too many. It is not viable to apply a multicriteria decision-making method with a so high number of alternatives, the reason why we will reduce the ordering alternatives to a few.

For the task of services restoration, the ordering alternatives that have been chosen (between the Decision-Maker and the analyst of the problem) are the following ones:

- A. Greater importance to the critical level. The services with higher critical level must be restored first. No matter which location is the service.
- B. Locations with more affected services, without ordering inside the location. This means that services belonging to a location with a lot of unavailable services will be restored first.
- C. Locations with more affected services, but ordering services inside the location. It is the same that alternative B, but inside a location critical services will be restored first. With the same critical level, services will be sorted by customer category. With the same category they will be sorted by bit rate.
- D. Locations with more critical affected services, and ordering inside the location. This means that services belonging to a location with a lot of unavailable critical services will be restored first. And, inside the location, the services will be ordered by critical level, then by customer category and then by bit rate.
- E. Locations with more unavailable services affected by some customer complaint, and ordering inside the location. This means that services belonging to a location with a lot of unavailable critical services will not be restored first if there is no complaints there. Once a location is chosen, the services will be ordered by the time its complaint is set: services with former active complaints must be restored first.

Full details of each one of these alternatives are not described here. As a general rule, we will consider that whenever a situation of "equality" in the evaluation of two services is reached, it will be more important the one that has higher bit rate, or has active complaints, or is critical (always ordered by level), or has a customer of higher category.

6 OPERATING PROCEDURE

Every hour, the amount of unavailable customer services along the networks is different. Network fails may be frequent, and within the same day we can have several important fails.

Depending on the affected services that exists at the moment of application of the Algorithm, the best "alternative" for that concrete application may not agree with the best alternative at another moment of the day.

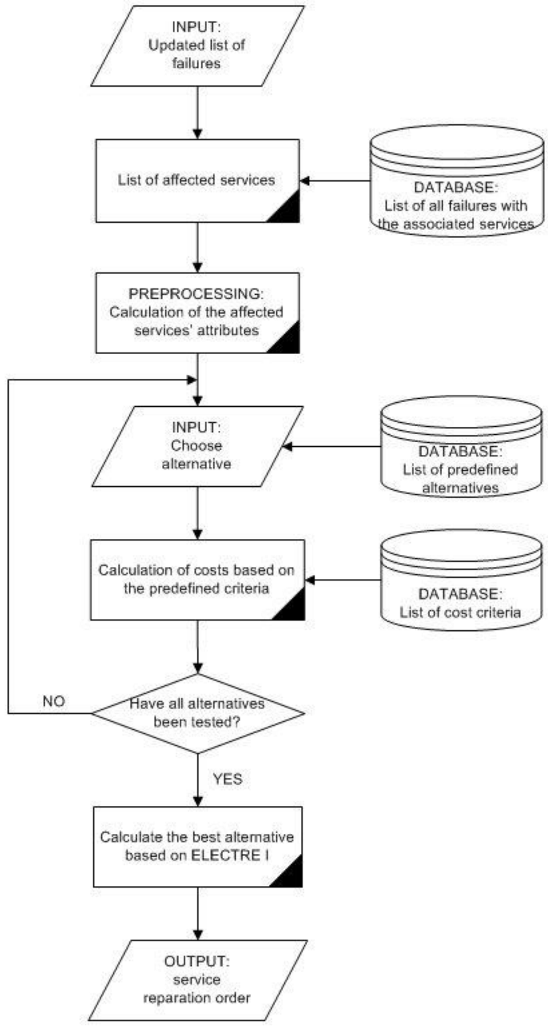


Figure 1: Operating procedure.

Therefore, the followings are the conditions of application of the algorithm:

- The decision to run the Algorithm will be taken by a maintenance operator. Normally this will happen when the operator is informed of a failure in the network. If the Algorithm has been applied once at the beginning of the day, but the operator is informed on a new network failure that has taken place later, the operator can decide to run the algorithm again (normally he will always do it).

- The Algorithm needs to take as input data the identification from all the services that are to be ordered. That is to say, before a network failure happens, it is needed to obtain the list of all the affected services. In addition, for each service not only its identification is needed, but also all the set of attributes that participates in the calculations. If throughout the day a second failure takes place, and the operator decides to run the algorithm again, it will be needed to have the list of affected services of the first and the second failure simultaneously.
- It is necessary to apply the five identified ordering alternatives (A, B, C, D and E) over all those services, so that five ordered lists of services are obtained.
- The result of ELECTRE I Algorithm, is a complete solution of arrangement of all the services. So, before knowing the winning ordering alternative, it is necessary to show to the user the corresponding list of ordered services. This is the final target of all this problem. This list will be used by the maintenance workers to begin the restoration process.

7 CONCLUSIONS

The algorithm explained here has been applied to several sets of real services, obtaining good results. It has been tested with data belonging to spanish telecommunications companies and to some South American companies. The computation times are acceptable even when the number of affected services are more than a thousand (what is normal in the real world).

Nevertheless, this is not the end of our investigation. One limitation of this algorithm is that the ordering alternatives have to be defined first. So we will keep on searching new algorithms (Aguezzoul, 2006) (Tiefeng, 2005) (Andina, 2006) to face the problem described in this paper.

REFERENCES

Romero, C., 1993. “Teoría de la decisión multicriterio: conceptos, técnicas y aplicaciones”.

Roy, B., 1991. “The Outranking Approach and the Foundations of ELECTRE Methods. Theory and Decision”.

Hayashi, M.; Fukuda, K.; Abe, T., 1996. "Evaluation of post-disaster restoration policies for telecommunication services”. IEEE Network Operations and Management Symposium, vol. 1, pp. 248 – 257.

ITU-T standard G.805, 2000. "Generic functional architecture of transport networks”.

M. D’Arienzo, A. Pescape, G. Ventre, 2004. "Dynamic Service Management in Heterogeneous Networks," *Journal of Network and Systems Management*, Vol. 12, No. 3.

Cameron A.F.; Thorne D.J.; Foster K.T.; Fisher S.I., 2004. "Fixed access network technologies," *BT Technology Journal*, Vol. 22, No. 2, pp. 48-59(12)

Yong-Kyung Lee; Dongmyun Lee, 2003. "Broadband access in Korea: experience and future perspective," *Communications Magazine, IEEE* , vol.41, no.12, pp. 30-36.

Antón, J.M. and Grau, J. B., 2004. “Madrid-Valencia high-speed rail line: a route selection”, Proceeding of the Institution of Civil Engineers. Transport 157, pp 153-161.

Antón, J.M.; Grau, J. B.; Andina, D., 2004. "Electre and AHP MCDM Methods Versus CP Method and the Official choice Applied to High-Speed Railway Layout alternative election". WSEAS Transactions on Business and Economics, vol. 1, pp. 64 – 70.

Aguezzoul, A. Rabenasolo, B. Jolly-Desodt, A.-M., 2006. "Multicriteria Decision Aid Tool for Third-Party Logistics Providers' Selection". International Conference on Service Systems and Service Management, vol. 2, pp. 912 – 916.

Tiefeng Zhang, Jinsha Yuan., 2005. "Decision-aid for power distribution system planning problems using ELECTRE III". The 7th International Power Engineering Conference, vol. 2, pp. 317.

Andina, D.; Vega-Corona, A.; Seijas, J.I.; Alarcón, M.J., 2006. "Application of Neural Networks". Computational Intelligence for Engineering and Manufacturing. Springer.

POTENTIAL AND CHALLENGES IN THE CONVERGING ICT AND MEDIA INDUSTRY

FITCE 2010 Proceedings

Jarmo Harno

Wipro Technologies, Finland

jarmo.harno@wipro.com

Keywords: Convergence, ICT and media, business architecture, value network, competitive strategies.

Abstract: The ICT and media industries are overcoming a fundamental change process as the means to access the information, entertainment and communication facilities are converging; the services are becoming more and more versatile, and new value networks of business players are formed. The change has several dimensions: technological, business transformational and in relation to the new ways of life. The presentation will summarize the requirements and potential of the traditional telecom players in the convergence of ICT and media industry. It will investigate the technology trends and business aspects contributing to the change, as well as the new role of the user.

1 INTRODUCTION

The ICT and media industries are overcoming a fundamental change process as the means to access the information, entertainment and communication facilities are converging; the services are becoming more and more versatile, and new value networks of business players are formed. The change has several dimensions: technological, business transformational and in relation to the new ways of life.

As the wireless communication capacities have grown, many usage patterns earlier tied with fixed access are now possible through small portable devices. Increasing amount of functionality can be packed to handhelds and the transmission capacities give access to all possible content and processing resources. Service orientation trend in ICT industry is gradually improving the processes and resource management to the level that makes it possible to cost efficiently provide end-user services that are easy to use and adaptable to the varying needs.

However, technical advancements are not enough for describing the change in the ICT and media field. As different capabilities are needed to be combined in providing versatile services through complicated mix of technologies, no single company can suppose to flourish without excellent partnering network. The players are adopting new roles in creating unique value that can give them dominant position in the end-to-end value chain. The orchestration of winning business architecture requires innovative thinking and functional processes crossing the corporate borders. The converging services will hide under their easy accessibility and usage all the technical subtlety and the multistage value networks. The user is the king who should be given freedom to create own usage scenarios, against payment that matches with the experienced benefit. Manifold services and products are combined and the virtual host of the user may generate revenues also from third parties through advertisements as connecting the user and the seller.

As the high speed access to resources through Internet is becoming a commodity and the customer data can be collected in intelligent and automated way, the telecommunications operators are facing increasing competition in customer ownership from other players in the field. Although network resources are big investment and a valuable asset, that capacity may face inflation in the long run, if not deployed wisely. With new technologies, the capacity growth can lead to oversupply in some scenarios, like happened with the telecom transport capacity in the previous decade, although currently it seems that the new services will easily consume at least all the provided mobile capacity .

For the tightening competition and increasingly complicated system and process management tasks, both new IT approaches and savings in operational costs are required. Organizations need to define their key assets and strengths, drive them to excellence and buy other resources as services from those players that best support their position in the value network. As the future is less and less dictated from company strongholds, shift from plain fixed plans to option planning is needed. Not all of the investments have to become profitable, but some of them may realize as growth options as unforeseeable usage scenarios grow up.

2 Technology integration through ubiquitous connectivity

There used to be clear separation between the sectors of the ICT field. Telephones were rather simple devices, the complexity was in the switching systems, and only the radio integration brought in some challenges. Computers, SW applications and data networks, on the other hand, had their own evolution path. Radio, TV and other electronic entertainment and information sector formed yet another ecosystem.

As the capacities of the mobile devices grew, the fundamentals of connectivity to all the ICT and media services are changing, meaning a huge business renewal. Many usage patterns earlier restricted to special equipment are now possible through a single mobile device. The end-to-end processes are developing to support more and more services, applications and content delivery through this channel.

The mobile access to vast amount of services builds on the preceding development of Internet as the World Wide Web. Service orientation and web services trend in the ICT industry is making it possible to cost efficiently provide end-user services and content easy to use and adaptable to the varying needs.

2.1 Mobile Access

A revolution, compared even with the industrial revolution, was touted during the biggest hype in the shift of the millennium relating to the advent of the third generation mobile systems. The whole new realm of services was seen to appear with the fast mobile data networks – and change the human life fundamentally. But already in 2001 these dreams turned out to be a phantom of a possible distant future. The change was not going to be that fast.

However, gradually the new networks were built, smart devices were launched, and lastly the business infrastructure and really usable services are being deployed. The world is starting to be ready for ICT and media provisioning to migrate to the mobile access. Increasing amount of functionality can be packed to handhelds, and the transmission capacities give access to the content and processing resources in the network with personalization and security. This is not to say that mobile would be the only access means, but that mobile access could be integrated to almost all electronic services.

2.2 Internet

The Internet has paved the way towards integration of all data and media sources and computing resources. Evolution of the service orientation in IT field and open source SW development are relying on the Internet and boosting its usage. At the same time technologies for users to create and share content are raising the Internet traffic into new heights.

As the Internet with all its resources is being integrated into the mobile personal computer, the new usage scenarios are gradually emerging –relating both to the new consumer habits and the enhanced business processes.

2.3 Service orientation

Provisioning of mobile services is much more complicated process than just building network capacity, as turned out in the beginning of the millennium. The user interface has to be very convenient, easy to learn and without disturbing lags, with reliable and visible charging - preferable with no charge at all, or even facilitating cost savings.

This sets high requirements not only for the devices, but especially for the service back-end and front-end, and most importantly to the business models and the whole business ecosystem to provide the services.

The operators' first attempts to master the service usage by their walled garden portals did not lead to good results. It proved to be unrealistic both from the provisioning and from the revenue perspective. Much had to change in the service development, provisioning systems and applications. Service oriented IT architectures were needed to support the service provisioning cost-effectively and collaboratively in the distributed Internet environment.

3 business architecture transformation

In the analysis of the converging ICT and media industry, the perspective of the technology evolution is not enough. We need to analyse the transformation of the whole business ecosystem. The provisioning of advanced services is not a single business actor endeavour, but a successful business case is dependent on co-operation of a partnering network, and even competing actors can support the business ecosystem as a whole.

As we have the developing technology architecture, we have also the evolving business architecture that utilises the technology and directs its development. Business architecture is not understood here to be limited inside the company, but rather to cover the whole value network.

Actually no business actor or technology developer can really direct the future usage scenarios, but they can deploy an architecture that is adaptive to the emerging usage needs.

3.1 Convergent ICT and Media Value Network

In the converging ICT and media business ecosystem there are usually several actors that have to play together to provide a winning service. There are particular roles in the service provisioning process, and relationships (interfaces) between them. The business actors try to capture the roles that they see most vital in their business plan and where they thus want to pursue strategic strengths.

The roles and relationships in the converging ICT and media value network were analysed in the European CELTIC framework project ECOSYS (2003-2007), Deliverable 03 (2004), and based on that work are illustrated in Figure 1. The respective roles are described in Table 1, and the relationships in Table 2.

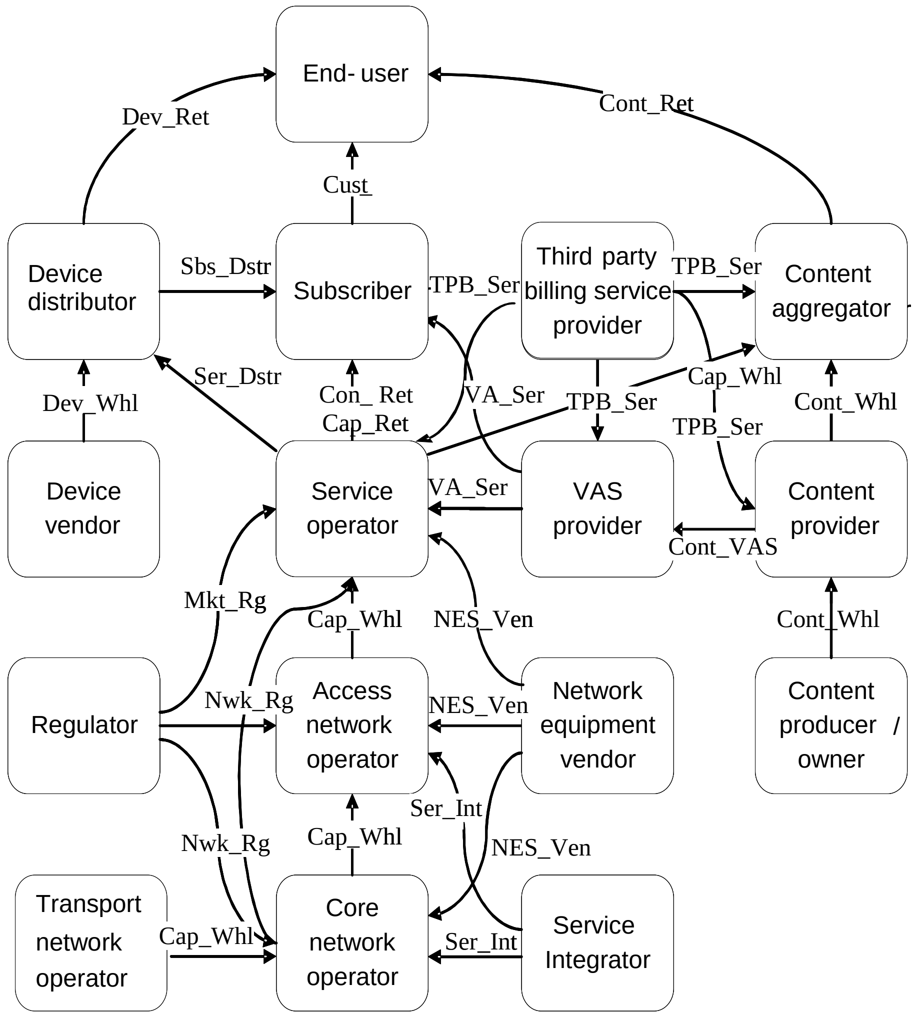


Figure 1: Roles and Relationships.

Table 4: Role definitions.

Role	Description
End-user	An entity taking up this role is the ultimate consumer of services and content provided in the communications ecosystem.
Subscriber	An entity taking up this role pays for the services and content provided in the communications ecosystem and has a direct relationship with the provider. However, this entity may not necessarily be the end-user of these services and content. A subscriber can be classified as: 1) Post-paid or pre-paid (based on mode of payment), 2) Consumer or business (based on type of entity)
Service operator	An entity taking up this role provides subscriptions for communication and distribution services over communications networks (fixed-line, mobile, terrestrial wireless, satellite). A Service operator does not own the networks. It enters into service level agreements (SLAs) with access and core network operators for network capacity. Other major responsibilities of this role include: 1) management of subscriber profiles, subscriber acquisition and retention 2) providing security services 3) charging and billing the subscribers for service usage. The Service operator may resell its network capacity to third parties such as content aggregators.
Content aggregator	An entity taking up this role acts as an intermediary between Content providers and End-users or Service operators for offering a portfolio of content (such as applications supporting e.g. social networking, games, music, TV programming or just ringtones and logos). It takes the content produced by Content providers, and converts it into a suitable format depending on the context (device capabilities, location, personal indications/preferences, etc.) and may rely on the VAS providers. It may offer several types of content at a single stop (e.g. a content portal) combining music, videos, books, etc. Content and applications may also be developed and provided for free, based on Open Source and/or revenues generated from advertisements, or increased usage value for the end-user, thus adding value to the subscription or device.
Content provider	An entity taking up this role publishes and sells content or applications developed itself or by other content producers. Other activities include marketing the content, conducting market research, etc. Sega, a game publisher is an example of a content provider.
Content producer/owner	An entity (business or individual) that takes up this role would develop and maintain content or applications such as music, games, etc. It may own the content or act as developer for an owner.
Value-added service provider (VASP)	An entity taking up this role provides services that are complementary in nature or add value to the basic set of services provided by a Service operator. Examples of such services include location-based services and presence.
Third party billing service provider	An entity taking up this role provides billing services to the Service operator, VASP or content aggregators. This role acts as a financial intermediary (or clearing house) between two or more Service operators. In return, it gets a share of the revenue earned by the operators. Credit card companies and banks fall in this category.
Device distributor	An entity taking up this role acts as the retailer for end-user device and related appliances.
Device vendor	An entity taking up this role provides customer equipment, such as desktop computers or mobile devices used by the end-users. It may also provide auxiliary software and hardware such as operating systems and batteries essential for the proper functioning of the device.
Access network operator	An entity taking up this role owns and administrates the access networks such as DSL, mobile, cable, broadcasting network, and provides transmission or bearer services to the Service operator. It sells access network capacity to Service operator enabling it to reach its subscribers.
Core network operator	An entity with this role owns and administrates the core network elements such as switches, routers, gateways and management units, offering core network capacity to the Service operator.
Transport network operator	An entity taking up this role owns and administrates the transport networks such as optical backbone networks and provides transport services to the Core network operator.
Network equipment vendor	An entity taking up this role primarily manufactures either by themselves or through original equipment manufacturers (OEMs) network elements and related services and distributes them to network operators (both access and core) as well as to Service operators.
Regulator	Regulator’s primary role is to maximise social welfare by suggesting and enforcing rules for sustaining a competitive market and technological environment. Regulator’s aim is to prohibit monopoly, misuse of dominant market power, and vertical integration where part of the value chain is subsidised to block competition. This role is important for formation of value networks.
Service integrator	An entity taking up this role provides service platform functionality that enables roles without extensive knowledge of the underlying communications system to offer services and application access to subscribers. Some examples of these functionalities include Parlay/OSA Gateway that provides APIs for charging and billing service, QoS etc.

Table 5: Relationship interface definitions.

Interface	Description
Cust_	This interface represents the relationship between the End-user and the Subscriber roles. Subscription may also have been made by an actor other than the end-user, e.g. by an employer or parents.
Dev_Ret	This interface represents the retail relationship developed between the Device distributor and End-user roles for the distribution of devices and related appliances.
Dev_Whl	This interface represents the relationship for the wholesale distribution of devices and related appliances. This relationship exists between the Device vendor and Device distributor.
Ser_Dstr	This interface represents the relationship developed between Service operator and distributor roles. Service operator may subsidise devices and/or pay commission to the distributor who facilitates the subscription between the Service operator and the device purchaser.
Sbs_Dstr	This interface represents the relationship developed between the Subscriber and Device distributor roles. The distributor may market the subscription with a Service operator in connection to the sold devices.
Con_Ret	This interface represents the relationship in making and fulfilling the retail contracts (subscriptions) for communications service packages, for voice, messaging, data transfer, media channels, etc. This relationship exists between Service operator and Subscriber.
Cap_Ret	This interface represents the relationship for the retail provisioning of transmission or communication service capacity, or transactions. This relationship exists between the Service operator and Subscriber.
Cont_Ret	This interface represents the relationship for the retail provisioning of content or applications. This relationship exists between: 1) Content aggregator and End-user/Subscriber 2) Content provider and End-user.
Cont_Whl	This interface represents the relationship for the wholesale provisioning of content. This relationship exists between: 1) Content aggregator and Service operator 2) Content provider and Service operator 3) Content provider and Content aggregator 4) Content producer and Content provider. (The first two relate to the case, where the Content aggregator/provider is a subcontractor of the Service operator; arrows are not depicted in Figure 1).
Cont_VAS	This interface represents the relationship between roles for the VAS provisioning and VAS related content and application provisioning, e.g. provisioning electronic maps, location related information and navigation SW, or TV-programming.
Cap_Whl	This interface represents the relationship between roles for the wholesale provisioning of network capability and capacity services, including accessibility, connectivity, security, QoS and charging. This relationship exists between: 1) Access network operator and Service operator 2) Core network operator and Access network operator 3) Core network operator and Service operator.
TPB_Ser	This interface represents the relationship between roles for the provisioning of third party billing services. This relationship exists between a third party billing provider and the following roles: 1) Service operator 2) VASP 3) Content aggregator/provider/producer.
Mkt_Rg	This interface represents the relationship developed between the regulator and Service operator roles for the implementation of market regulations in order to provide market fairness. Pricing regulation is one example of such market regulation.
Nwk_Rg	This interface represents the relationship that exists between regulator and the Network operator roles (both access and core) for the implementation of network regulations such as mobile spectrum distribution, wholesale pricing and interconnection.
Ser_Int	This interface represents the relationship that exists between the service integrator and Network operator (both access and core) roles for the implementation of service platforms that enable network independent services and applications.
VA_Ser	This interface represents the relationship between a Value-added service provider and the role of: 1) Service operator 2) (direct) Subscriber.
NES_Ven	This interface represents Network equipment vendor’s relationship with the following roles: 1) Core network operator 2) Access network operator 3) Service operator.

3.2 Competing Actors in the New ICT and Media Business

The business to provide new services through mobile networks was first seen to be dominated by the telecommunications operators. The i-mode in Japan was seen as a model for operator controlled ecosystem also by European operators in the advent of 3G data services, but that model did not really flourish Europe.

Gradually the operators' walled gardens started to be opened, but the partnering in creation of new services did not really bear much fruit. Even until recent times no real success among the services emerged after the SMS.

Towards the end of the last decade the mobile device vendors started to shift their focus towards wider end-to-end solution provisioning, where the services would play an important role. Nokia launched their OVI service, but only when Apple presented their iPhone concept the control of the value chain appeared in a new constellation. Apple, partnering with chosen operators, not only provided the handheld device with Internet oriented architecture, but also content and applications through the iTunes, and had even control on the related long term subscription.

From another frontier, Google with their partners have developed the Android devices, again with the full Internet based end-to-end value chain, combining the successful advertisement based business model into the mobile service provisioning.

The communications operators, broadcasters and other content providers have to consider their strategies, together with system integrators, in the evolving ecosystem.

4 Competitive strategies

Transformation in several dimensions is needed for a telecommunications operator to be competitive in the converging ICT and media industry.

4.1 Innovation and Value Capturing

The ICT and media business ecosystem is getting more and more complicated, novel value chains with new players or positions are emerging rapidly, making dramatic changes possible, even expectable, in the business environment of the traditional players.

Technology development is one dimension in the disruption that can take place, but even more there is question of a change in the usage patterns. Some legacy features in a usage context may suddenly turn out to be unimportant and some new aspects to be of greatest interest.

As a current example, the telephone call and messaging features in the mobile phones, being mainly continuities from the fixed phones, were suddenly out of the focus of the most advanced users and that was spreading rapidly to the wider user population. In the Apple's iPhone the easy usage of Internet and versatile applications by the touch screen, overrode the shortages in some calling features. The whole user setting was transformed - finally the endless world of computer applications was brought into the mobile device in a lucrative way.

Same kind of paradigm change can happen in many points of the new complicated business architecture of the ICT and media value network. Traditionally strong players may suddenly lose their advantage and control position in the business, if more innovative players can bring in some new user concepts that make the assets of the incumbent players practically almost worthless. "Linear" development on the current, however strong, track, or cost reductions and competition only with price, is not a safe strategy anymore.

Creativity is indispensable in preparing for the future. It requires scouting for the possible new scenarios, developing new competencies, products and relationships to be ready to serve the new usage patterns that emerge, and capture the value in there.

The common situation during the last decade in the telecommunications field has been the technology push and market reluctance that has led in many cases to competition by price. A market disruption and break-through with growing data service revenues required much more than most of the technocrats anticipated. Technology is very important factor in the economical development and technology strategy has been vital for the survival in the competition. It has become currently even more complicated, as a system of interwoven technology and business architectures or ecosystems. Changes in one part of the ecosystem have effects throughout the system and the totality has to be kept healthy.

But the human dimension adds the most unpredictable component. It is not easy to get people to adopt advanced products as they usually involve trade-offs related to their old patterns of action and require some change. Against common assumption people do not require necessary simplicity, but they have to be able to see the benefits clearly and worthwhile, relating to their authentic needs.

It is not enough to find out the points of unsatisfied demand for functionality or service to create value, but it is also needed to be able to build the profitable value creation machinery so that others cannot easily copy it. The value capturing requires unique resources, channels or brand; or the unique knowledge may possibly be protected through patents and copyrights, or it is so much dispersed in the organisation that it simply cannot be copied.

Despite the importance of the technology development, the actual breakthrough of new service paradigm is most probably not foreseeable from the sole technology perspective, but after the period of technology advancement the total usage setting is suddenly ready for the new behaviour patterns to rise, first among the advanced user group and starting to diffuse to the wider user population.

People that are close to the service provisioning, understanding the user community and its needs, are in a better position to create successful products and services than those with the technology push approach. Recently, the participatory role of the user in the new innovations has got more attention. As the information technology is reaching the level where the infrastructure allows rather easy development and large scale integration of applications and content, the emphasis is shifting from the technology itself to the actual human needs and the process demands in the businesses.

4.2 New Business Models and Position in the Value Network

As looking at Figure 1, there are several roles in the value network that the current and emerging players can take. Many of those roles are in new deal in the shake-up of the industry.

For example, the device vendors, Apple most successfully, have adopted new roles and moved towards controlling the whole end-to-end value chain. Especially the Content aggregator and Value-added service provider roles have risen up in the importance.

As mentioned, Internet is more and more the channel for distributing services, content and applications. The old world of special protocols for different services is disappearing and generic high usability web interfaces are taking over. The small mobile devices have become capable to provide the user interface to almost any content and application.

User profile information is not anymore necessarily only in the hands of some Service operators through subscriptions, but residing in the social networks of the Internet - and the connections to the Internet are versatile and not in the hands of any single player. This makes the services and content itself the king. The users float to the channels that provide best value for the money. And buy devices and use services that they like most and that are affordable.

If some service totality, including device, connections, services and applications available, is more lucrative than other propositions, the competitors are in trouble whatever their position in the value network used to be. Everyone wants to be in the winning team, but the one who owns the key innovations sets the rules for the other partners.

What will be the next big thing in the ICT and media market? The key feature may be in the device or in the network/server side, but most probably something that combines elements from both realms.

The current transformation in usage is mostly taking place in the consumer services and applications (although business users are utilizing the same services as well). The next wave may be in the revolution of the business services, as the corporate systems are being integrated to the all-pervasive Internet media. This integration would be both internal and external for the company, both private and public. Presence and connections through the converging communication and media channels is becoming a must in the new economy of ubiquitous virtual presence.

Surely not all players should rush to the Content/application aggregator or VAS provider role, but all elements of the value network depicted in Figure 1 are needed. Regulatory trend is clearly to break up vertical integration, so that on each level of the value chain the players have to compete for being the best in the breed, even when being part of a larger corporation.

In the efficient market conditions the return on invested capital should not be dependent just on the position in the value chain – the competition should take care of that, but if a player has some intellectual or physical property that cannot be rapidly copied or rivalled, it can make good profit in any role.

In the rapidly changing complicated business situation the challenge is not only to see the technical development, but especially the needed changes to the whole of the business and technical architecture. New service paradigm may require completely novel technology approach, combined with new thinking in the business relationships, revenue streams and sharing, sales channels, etc. The change may be so radical that a start from a new foundation is needed for becoming able to adapt with the emerging reality, as the legacy approach cannot really be tuned optimal for the new requirements.

It is typical in the technology business disruption that new attributes start to be crucial in the usage experience and thus competition, as the old emphasized parameters start to be partly self evident, provided by every competitor, and partly evolution that is not really needed in the new usage setting. For technology disruptions, see Adner (2002) and Christensen (1992).

4.3 Competitive Assets in the New Business

Operators' core competence has been traditionally the efficient running of the networks. Together with customer management, networks have been the key asset of them. The optimized deployment of the new technology in the networks continues to be important for the operator success, but due to the more and more complex services and related value networks (business architecture) the front-end and back-end business support systems are requiring increasing focus.

On the network side, the utilization of the current networks and future investments should be optimized. Well-grounded technology strategy is needed, where convergence gives new possibilities for optimal usage of the network resources. But most benefits are foreseen in the total renewal of the organisation structure and business processes to benefit from the synergy in providing services seamlessly utilizing all networks, whether fixed or mobile.

A recent study, based on the ECOSYS (2003-2007) project (Harno et al. 2009) demonstrates that the yearly OPEX savings relating directly to the converged service provisioning are substantial, growing to over 800 M€ yearly savings, being about 7% of the total yearly OPEX of the modeled integrated (fixed-mobile) operator, which has its total OPEX at the level of 12 B€ (see Figure 2).

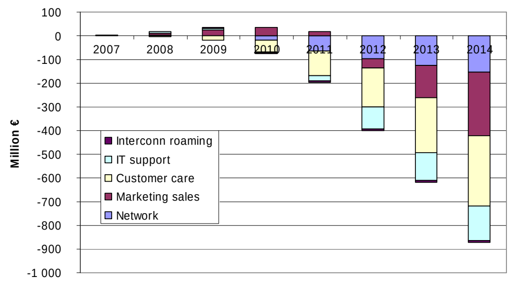


Figure 2: FMC case OPEX (delta) breakdown (negative values show OPEX savings)

From these results, we can see that after some years, the operator with fixed-mobile convergence (FMC) approach will reduce its OPEX compared to one without convergence. The trend in the OPEX savings is growing even after the study period, giving a considerable competitive advantage for the integrated operator selecting the FMC path.

Major contributors to overall OPEX savings over the entire study period, are customer care (40%), network maintenance and management (19%), IT support (15%), and expenditures for marketing, sales, and customer acquisition (12%). Figures for marketing will increase in the beginning as new products and services have to be established. The saving increases towards end of the study period because a concentration to converged products will occur. Similarly, the customer care effort will be reduced because of less separate subscriptions within the integrated converged network compared to two separate networks and multiple subscriptions per person.

The effects on the operator business are not limited to the OPEX savings, but the new convergence services bring new revenue streams and especially increase the possibilities to cross-sell and retain the customers in the tightening competition. Their contribution to the net result is suggested to be even greater than the OPEX savings according to the referred study, even when the cannibalization effects due to the convergent service packaging are taken into account.

Hand in hand with the advanced technologies and convergence, the legacy telecom business as a separate industry is transforming into integral part of the wider ICT and media industry. New innovative services require figuring out the new business architecture, understood broadly as including the full value network with all partnering players needed in the service ecosystem.

Mobile device vendors have mainly taken the initiative in building the ecosystems from content and application producers to service provisioning, seeing this crucial for selling the devices as packaged with easy to use services. Google is providing an ultimate model, having the access networks and devices as secondary pieces of the whole, the core asset being in the Internet - its servers and service intelligence sold mainly to the advertisers, not to the consumers.

How can the traditional operators compete in this changing environment? The creativity has already been emphasized. The operators should utilize the assets not only efficiently but also innovatively, to compete or co-operate with the Internet players. The time of the “walled gardens” is over and due to the regulation it is hard to hinder the Internet companies to strain the access network with their service provisioning.

This is deepening the broadband incentive problem and bringing it also into the mobile access. Intelligent pricing schemes are needed to prohibit the network to be overloaded by traffic unprofitable for the network operator. The average user likes predictability of the bill, and thus the flat rate is preferred, but it can include certain limitations for example in the form of quotas (see Harno 2010). By segmentation, traffic restrictions related to time of day or location combined with price reductions for the low-end users, and quality of service provisioning to the high-end customers and services, can provide win-win situations in the value chain of the service provider – network provider – end customer.

Partnering has been chosen by many operators nowadays as being a better and more profitable way in providing good choice of services to the consumers than “walled gardens” and access restrictions (see Harno 2010). By right pricing, even the VoIP services have not become a threat for the operators.

Even though telecom operators have not been strong in providing managed services to their business customers, in the area where the big system integrators have dominated, the setting is changing in relation to some of the new trends. For instance cloud computing is a very network intensive service and network operators can exceptionally well build and run the data centers hosting the cloud services. Many kind of services can be hosted by the operator and it can provide platforms to other players for easy implementation of their services utilizing its advantage in the integration of the servers to the network.

There are also many verticals, e.g. health care, construction etc., where there is increasing need for fast transmission of big amounts of data, even with mobility, for remote diagnostics, consultation etc. Special requirements are set for the quality of service, which can be best provided by the network operators. Management of the mobile terminal base, applications, and access to the company resources, with all the security measures, can be provided by the operator too.

Service provisioning in the new data communications era requires partnering, not only with aforementioned players in the communications and media value network, but also with subcontractors in business transformation and implementation of the new processes and services. As mentioned, the attention is shifting from the network operation, which the operators traditionally master well and which is more and more outsourced, to the front-end and back-end BSS systems, service creation machinery, and general IT architecture. The service-oriented (SOA) thinking has been the major trend in information technology during the first decade of 2000 and has been adopted to the communications industry especially through the Tele Management (TM) Forum (2010). The TM Forum Solution Frameworks (NGOSS) can be viewed as master plans to guide towards service-orientation, by implementation of the processes by reusable Business Services. Utilization of COTS products and open source SW instead of bespoke applications and solutions with expensive licenses can also be seen as a growing orientation, as well as outsourcing large parts of the development and maintenance to the system integrators to gain costs savings and better return on investments. Business intelligence is one of the key focuses in the new communications and media industry. As the converging whole is growing more and more complicated, running the business requires combining information from plenty of sources and distributing it to multiple directions, but maintaining at the same time a unified interface towards each customer – and do that fast enough to keep up in the competition.

It is not anymore a question of understanding the market opportunities and developing a plan to be executed over the next period of time. In the current very dynamic business environment decisions have to be done fast but creatively as the underlying complex ecosystem is changing constantly and the changes are interacting with each other making sudden disruptions possible.

After developing systems for dedicated solutions for customers (CRM), suppliers (SCM), etc. and separate system yet for the senior management (ERP), it has been hard to integrate data in all these different systems, whether through centralized repository or some middleware. Web-interface based service access has increased popularity within the SOA approach and recently the enterprise mashup concept has risen, offering a new way to accelerate the process of bringing together information throughout organizations by using Web data standards to access critical information. This approach even empowers business users to construct reports that utilize real-time information sources in driving the organization. For example in the campaign management it should be possible to segment customers according to all their converged access devices, service usage patterns and customer history, over all the provisioning channels.

5 CONCLUSIONS

The current competitive situation in the ICT and media field calls attention on the need for strategic considerations relating to innovation, new business models and partnering schemes, as well as transformation needed for the optimal utilization of own competitive assets in the new situation.

Preparing for the future requires efficient utilization of competitive information and agile business process implementation, but hyper dynamic environment requires, in addition to fast decisions, ability to take controlled risks.

As unforeseeable disruptions are potential, option thinking is needed in addition to incremental improvements. Start-ups from completely new technology and business network ground might be this kind of options, on which reasonable amount of money is invested and if the world goes to certain direction, some of them can be executed – i.e. new concept can be deployed further rapidly and new market captured. On the other hand, if the world goes to another direction, where other kinds of usage scenarios arise, particular options may turn out to be worthless, but some other may be executable instead.

REFERENCES

Adner, R.: When Are Technologies Disruptive: A Demand-Based View of the Emergence of Competition. *Strategic Management Journal*, Vol. 23(8), 2002, 667-688.

Christensen, C., M.: Exploring the limits of the technology S-curve. *Production and Operations management* 1(4), 1992, 334-366.

ECOSYS (2003-2007). Project web site: http://www.optcomm.di.uoa.gr/ecosys/deliverableslist.html.

ECOSYS Deliverable 3 (2004). Business models in telecommunications. Retrievable from: http://www.optcomm.di.uoa.gr/ecosys/deliverables/ECOSYS_Del03_v1.0.pdf.

Harno, J., Kumar, K.R.R., Heikkinen, M., Kind, M., Monath, T., Von Hugo, D., 2009. Service Offerings for Fixed-Mobile Convergence Scenario: An Integrated Operator Case. *International Journal of Business Data Communications and Networking (IJBDCN)*, Volume 5, Issue 3, pp. 1-16.

Harno, J., 2010. Impact of 3G and Beyond Technology Development and Pricing on Mobile Data Service Provisioning, Usage and Diffusion. *Telematics and Informatics*, Vol. 27, Issue 3, pp. 269–282.

Tele Management (TM) Forum (2010). Retrievable from: http://www.tmforum.org/.

CONVERGING TELECOM CLOUDS

Huib Ekkelenkamp
Atos Origin, Papendorpsweg 93, 3528 BJ Utrecht, The Netherlands
huib.ekkelenkamp@atosorigin.com

Keywords: Telecom, Convergence, Cloud Computing, Telco Service Delivery

Abstract: The Telecom environment is changing continuously through market and technological developments. In particular in Europe the growth of the telecom business shows saturation in number of customers and terminals sold. However, irrespective of the economic downturn the growth of telecom services and related traffic is steady. In particular data services like mobile internet and messaging grow faster than expected. Smart handsets, flat data rates, increased throughput and versatile user interfaces stimulate mobile data use. The supporting network technology has been standardised to a large extend, supporting all global markets with a harmonised infrastructure capable to deliver a wide range of services at high speed. As most Telco's base their future network development and roll-out on the same broadband standards, the sustainable competitive advantage obtained via the network is diminishing. The focus shifts to communication solutions and services which support directly the business and requirements of the users. The underlying network is taken for granted. The users want a wide range of services, ease of use, high performance and flexibility at low cost. Three developments have large influence: convergence, outsourcing / cloud computing and increasing the value chain. New technical developments for mobile and fixed networks are used to support broadband connectivity. Dedicated platforms for application download and Machine-2-Machine communication create new opportunities for Telco's. This will require new business models for IT support and participation in the value chain. It results in a new orientation of Telco's and System Integrators.

INTRODUCTION

The focus of the Telecom industry was in the period 2006-2008 on Next Generation Networks (NGN). In 2008-2009 every development was followed by 2.0 like Telco 2.0. Now in the year of Charles Darwin 2010 it seems that network technologies get Evolved: HSPA+ (HSPA Evolution) and Long Term Evolution (LTE) with Evolved Packet Services (EPS) and Evolved Packet Core (EPC).
To get some guidance in these Converging Telecom Clouds, the metaphor of Jacob's staff, an old navigation tool, can help us here in Santiago de Compostela to find our position. With one eye on the future demand horizon and one eye on the technology stars we should be able to predict some trends.
The total Telecom revenue in Europe in 2009 of about 350 b€, was 43% for fixed voice & data (consumer & business), 47% mobile voice and data (consumer & business) and 10% for TV.
Over 20% of mobile revenue is data (of which 4% is mobile internet) and 80% is voice. Mobile data traffic has already exceeded mobile voice traffic.

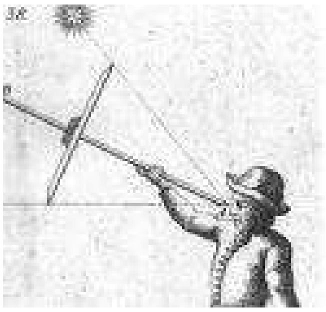


Figure 1 Jacob's staff navigates between the demand horizon and technology stars

Main developments are: increasing bandwidth (faster access to information), increasing processing power and performance (more high quality services) and increasing power efficiency (longer use). Telecom development will be affected by the following trends further elaborated hereafter:

> Convergence
> Cloud computing / outsourcing
> Increasing the value chain, social networks

Convergence everywhere

Converged service offering means fixed, mobile, flexible bandwidth, network and location independent provisioning of services and applications. Figure 2 shows some important convergence areas. With the bandwidth extension and capacity growth of mobile networks, differences between fixed and mobile access tend to become so small that the network is no distinguishing factor anymore.
Telco's should support with their services their business customers with strategic goals as growing market share, productivity partnerships and faster time to market. They should support with their solutions cost efficient operations with high levels of security.

This requires a customer-centric approach by access flexibility, service uniformity and unified communications. For Telco's this means bandwidth optimisation and network simplification (IP-based).

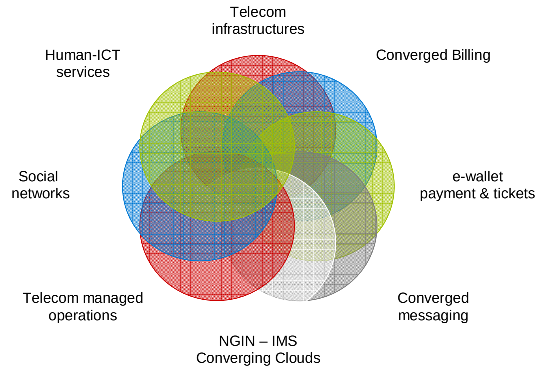


Figure 2 Convergence everywhere

The following trends determine the role of the Telco's:

a. Convergence of infrastructures, services and applications
b. Availability, versatility and price of mobile user terminals
c. Growth of the Intelligent web (read, write, execute, semantic)
d. Adaptation of outsourcing and Cloud computing (infrastructure / application / platform as a service)
e. Use of Ubiquitous computing (fast, grid)
f. Human-ICT convergence (decision support, nomadic use)
g. Evolution of social networks
h. Use of virtual worlds and artificial intelligence

Some of these trends have as effects:

> **Convergence:** raises quality of service and increases use and thus telecom traffic

> **Cloud computing and outsourcing:** decreases cost of ICT investment, increases demand for services

> **Intelligent web, virtual world and social networks:** stimulate innovative services and platforms

The trends are influenced by the following factors:

> Replacement of fixed by wireless technologies

> Ownership and control of infrastructure

> Effect and speed of standardisation

> Open versus closed technologies (security, proprietary, reliability, exclusivity aspects)

> Availability of capital for funding of infrastructure

> Stimulus by government through funding or regulation

> Effect of energy prices and environmental constrains

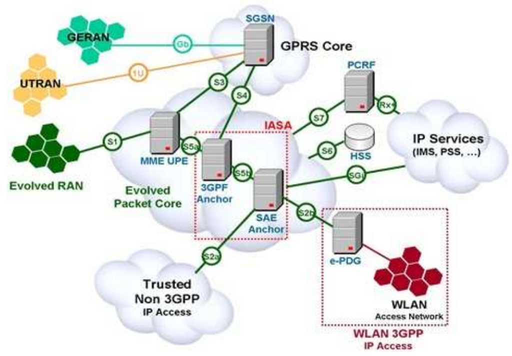
> Flexible and future proof solutions

> Use of embedded sensors and RFID

> Digital storage capability and decreasing cost

> Stimulating or restraining socio-economic environment

Mobile network developments have shown a faster growth in subscribers, services and traffic than fixed networks. In particular data traffic grows fast requiring new mobile access networks.
At this moment the last generation mobile LTE (Long Term Evolution) is introduced. This broadband mobile network is in particular designed for transport of data. Voice is supposed to be handled over IP. Interaction with existing 2G and 3G (circuit switched) mobile networks as well as with the fixed network will require some further adaptation of the 3GPP standards. It is expected that the standardisation progress will enable fast introduction of voice services over LTE in 2011. In particular the network handling capabilities of combined voice and data traffic (including the peak traffic loads) will lead to considerable cost savings in infrastructure investments. This All-IP network should be capable to handle all different services, originated in legacy or in the most advanced packet based sources.



LTE provides mobile broadband connectivity to 75 Mbit/s uplink and 300 Mbit/s downlink, depending on availability and capacity of the radio network. As shown in Figure 3, the Evolved Packet Core (EPC) is designed for seamless integration with IP-based networks. LTE has lower latency than 3G HSPA (High Speed Packet Access) and is capable to handle more channels or users per frequency band. This is a.o. attractive for gamers. The flexibility in the use of spectrum is a major advantage of LTE. Voice services will grow after voice standardisation and terminal availability for combined 2G/3G/LTE use has been solved. The IP Multimedia Subsystem (IMS) will enable the connectivity between different mobile and fixed networks. It will speed up the initial set-up of data connections. On top of HSPA, LTE gives Telco’s an increase of ARPU for data based services. The business case of LTE is therefore supported by increased revenue, due to increased usage like online gaming and video/TV, lower cost per unit of traffic and larger customer base. This will result (depending on network coverage) in a positive business case for Telco’s in Western Europe and the US. For network coverage there is an optimum to be determined by urban coverage, business support, heavy data users etc. LTE is based on shared and broadcast channels and is therefore a most efficient converged mobile communication system. It is often positioned as a further enhancement of the

mobile network enabling broadband mobile connectivity.
The next step for LTE evolution is LTE Advanced and is currently being standardized in 3GPP Release 10. LTE started in 2010 commercial in operational networks in Scandinavia with data based services. LTE handsets are expected to become available in Q1 2011. It is expected that further roll-out will follow from Q2 2011 onwards.
The availability of mobile frequency bands both in the network and the handset is a key element in the coverage and roll-out. Frequency bands differ between countries and could pose restrictions on international roaming.

Figure 3 LTE mobile broadband connected to existing networks

The growth of WiMAx is unclear at this moment. The mobile WiMAX, based on the 802.16e standard provides to 40 Mbit/s (theoretical) and with the IEEE 802.16m update expected to offer up to 1 Gbit/s fixed speeds. WiMAX can therefore be considered as a wireless broadband access technology complementary to LTE in niche markets.
The enormous growth of mobile broadband stimulated by the smart phones, in particular the Apple iPhone and Google Android phones has increased mobile data traffic considerable. This results in network congestion, leading to a lower quality of service. Mobile video calling might result in another overload of the network.
Flat rate has stimulated the use of mobile broadband as well. Telco’s are reconsidering their tariff strategy as one size fits all is not the right approach. Flat rates will be changed in pricing models which are personalised, flexible and dynamic. This means that customers will pay for what they use.
Video is the key driver for new mobile Internet business models. The possibility for users to select (temporarily) their own bandwidth and related tariff to get the desired video quality might result in a different revenue structure for the operator.

Fixed broadband on operator local loops is growing from ADSL2+ (now 12 Mbit/s) via VDSL2+ (now 24 Mbit/s on local pairs of 1 km from the exchange), to hybrid FttH with VDSL2+ and FttH (now 100 Mbit/s on new local connections).
For cable and Hybrid Fibre Coax (HFC) Networks DOCSIS 3.0 and HFC Next Generation provides 100 Mbit/s to 1 Gbit/s. Of all services video and High Definition TV with 8 Mbit/s per channel (minimal 2 channels per household) are dominating the bandwidth demand.
It is expected that the growth in Europe in data for consumers will be 50% per year on fixed connections and 100% per year on mobile connections. The average download rate demand for 2020 is expected to be at least 75 Mbit/s to 400 Mbit/s which requires optical fibre in the local loop. In general it can be stated that the fixed network can provide an order (x10) faster access than the mobile wireless network.
As answer to the cable TV companies, IP TV has become a serious alternative of Telco’s on their existing DSL networks. It provides an attractive enhancement of the services portfolio for the three main screens TV, PC and Smartphone. Here again, the required capacity will determine the used

access technology. For the fixed network this will lead to faster roll-out of fibre networks.
For mobile TV, LTE could provide the required bandwidth in the future. However, a separate TV overlay network like DVB-H is considered more efficient. Also 3GPP Multimedia Broadcast Multicast Services (MBMS) or the combination of DVB-H & MBMS is possible.
For user acceptance the following aspects are important: watch everywhere, affordable pricing, excellent quality, high-quality content, easy handling, no frills, low latency and click-trough access, multiple devices. Interactivity stimulates TV usage. For operators important aspects are spectrum availability, seamless combination of unicast and broadcast service.

Figure 4 IP TV for all screen formats

In The Netherlands a relative high demand for mobile and fixed communication is seen. The mobile telephony penetration is in 2010 about 130% of the population (50% prepaid, 50% post-paid) and the costs are the lowest in Europe. Data / Internet on fixed broadband connections are now covering 40% of the population (90% of households) with about 50% via DSL and 50% cable / fibre. TV reached 100% of the households (55% digital, 45% analogue). FttH reached 700,000 houses (10% of the households). Bundling is mainly with fixed telephony + internet + TV; mobile services are separate and seldom in bundles.

Clouds for Telecom, Iaas, PaaS, SaaS

Cloud Computing for Telco’s is still in its infancy. Most Telco’s have their own networks and supporting IT systems and are in the process of considering outsourcing of non-core parts.
The concept Cloud Computing incorporates combinations of the following:

- > Infrastructure as a Service (IaaS)
- > Platform as a Service (PaaS)
- > Software as a Service (SaaS)

Figure 5 Converging Clouds

- **Infrastructure as a Service (IaaS)**
IaaS is at the lowest layer and is a means of delivering basic storage and compute capabilities as standardized services over the network. Servers, storage systems, switches, routers, and other systems are pooled (through virtualization

technology, for example) to handle specific types of workloads — from batch processing to server/storage augmentation during peak loads.

- **Platform as a Service (PaaS)**
  The middle layer, or PaaS, is the encapsulation of a development environment abstraction and the packaging of a payload of services. PaaS offerings can provide for every phase of software development and testing, or they can be specialized around a particular area, such as content management.
- **Software as a Service (SaaS)**
  SaaS is at the highest layer and features a complete application offered as a service, on demand, via multi-tenancy — meaning a single instance of the software runs on the Cloud Computing provider's infrastructure and serves multiple client organizations. The most widely known example of SaaS is Salesforce.com, but there are now many others, including the Google Apps offering of basic business services such as e-mail.

Cloud computing is a paradigm of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet. Users don't need to have knowledge over the technology infrastructure in the "cloud“.

The first step for small to moderate size Telco's is virtualisation leading to IaaS. The subsequent step is using SaaS. This can start with office applications, testing services, database services and lead to gradual further transfer of BSS and OSS applications. Also the use of PaaS but than also by the Telco herself is very probable.

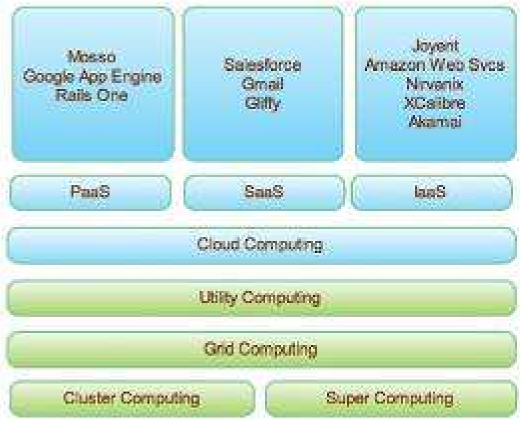


Figure 6 Cloud computing PaaS, SaaS and IaaS

New orientation, how to participate in the value chain

Telco's are reconsidering their core business. They focus on service provisioning mostly but not necessary via their own telecom networks. With a shift from fixed to mobile communication (or wireless access) new challenges emerge. Customer self support using web portals for ordering, complaint handling and billing reduce the number of Business Support Systems (BSS) and Operational Support Systems (OSS). Telco's realise that IT systems are crucial for efficient operations but see IT not as part of their core business. Even the Telecom network itself is nowadays not always considered as an asset they should own. This means network design, procurement and maintenance are than not considered as core activities of the Telco anymore but are shifted to or stayed with the network system supplier. The supplier of telecom equipment can manage his own network systems on behalf of the Telco. Here we see a shift of activities of telecom suppliers from hardware to services.

Telecom Network outsourcing becomes more attractive for Telco's because:

- Services are more complicated as more devices have to be supported
- Telecom networks heavily rely on IT elements both software and hardware and are therefore more related to the IT world than to the traditional Telecom world, requiring different knowledge and skills
- End-user devices or terminals are more complicated than ever and require much IT knowledge
- Networks are growing in complexity due to the large number of devices to be supported and the interworking with legacy networks and systems
- Standardisation is a major issue. The large number of standards produces by the large number of standardisation organisations requires considerable effort to follow and translate to operational environments.

New business models are required to support the new role and orientation in the Telecom sector. For example instead of SLA contracts based on performance and time spent, a fixed cost for IT support per subscriber per year (e.g. 5 €) could lead to a different approach. Other business models for innovation support in terms of revenue sharing and risk taking could also change the cooperation with Telco's.
Platform services support two-sided business models, acting to wholesale and communities (developers, retailers, government etc.) on the one side and users (and user segments) on the other side.

For the Telecom Sector the following business models and payment methods apply:

- Sales once or per separate unit (macro and micro payments)
- Advertising based payment (content can be free of charge)
- Pay per click (under user control)
- Revenue sharing (like Amazon and Google)
- Usage based payment (like mobile prepaid) or transaction based payment
- Subscription based (can be combined with advertising or usage based)
- Licensing (for applications and platforms)
- Free donations (like Wikipedia)

This flexibility will support partnerships for Telco's with IT companies and enable other positions in the value chain. As shown in Figure 7 this can be a linear or non-linear value chain between content providers, aggregators, application & platform providers, service providers and end users.

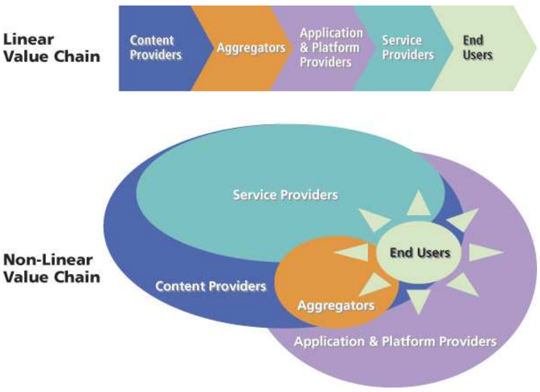


Figure 7 Telecom Linear and Non-Linear Value Chain

The IT support will change when different business models are used. Instead of implementing projects based on investments by Telco's, Systems Integrators might do the investments and let the Telco's pay for the usage. It is also a result from Cloud Computing where usage-based payment is a key element.

Apps stores, SDPs, M2M

The fast growing market for mobile smart phones is stimulated by the mobile applications which can be downloaded from Application Stores using Service Delivery Platforms. Mobile application downloads are expected to grow from 9 billion in 2009 to 50 billion in 2012, a year-on-year growth of over 90%. This is expected to become an additional revenue generator for Telco's. For Europe alone the total revenue in 2009 was 1 billion € and is estimated for 2012 on about 5 billion €. On a global scale these figures are 3 billion € and 12 billion € respectively. It is therefore expected that every Telco will assess the opportunities of SDPs and Application Stores. The main players are Telco's or Communication Service Providers (CSPs), device manufacturers, OS suppliers, and Apps Store platform providers. IT companies like Atos Origin have capabilities to support this development and generate new revenue in this innovative area. This can be the delivery or hosting of a platform (PaaS). It will require partnerships with suppliers.

A still moderate, but fast growing area is Machine-to-Machine (M2M) communication via mobile networks. Machines with built-in SIMs can exchange their status or instructions instantaneously with back-office systems. This will lead to large volumes in transaction-based communication.
In particular in healthcare the largest volumes and growth are expected. Here the key words are now e-health and m-health as most of these transactions are mobile-based. To develop this portfolio further, close cooperation between a Telco, an applications developer in the medical sector a systems integrator and a users group is required. Here also health insurance companies play a role.

Machine to Machine communication is expected to grow fast: with over 200 million mobile connected devices estimated in 2011 (Cisco even expects over a billion) and soaring revenues a large market potential is available. Figure 8 gives some focus areas, application groups and service sectors.

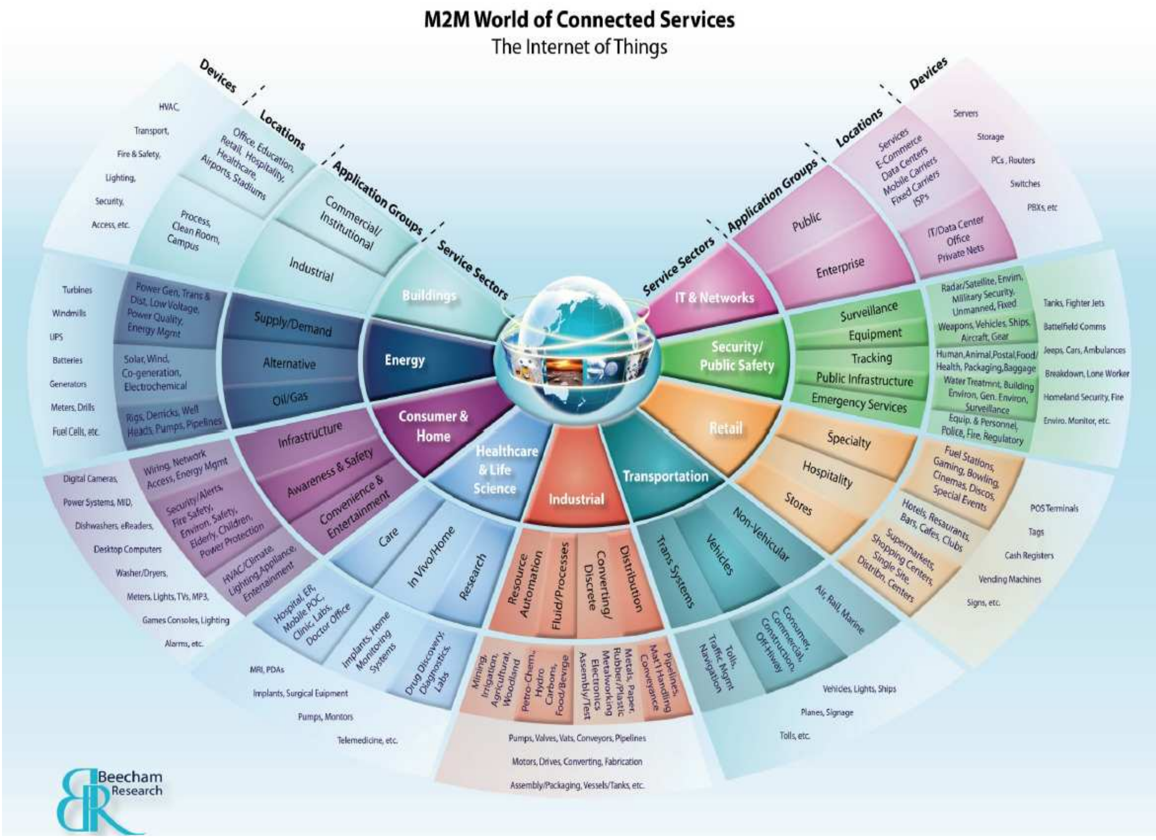


Figure 8 M2M market evolution

Changing IT support

The BSS/OSS market is now supported by a wide range of pre-integrated, out-of-the-box products. Application vendors have extended their position by providing also professional services like consulting, architectural guidance and system integration. The recent acquisitions of packages and suites by software vendors have extended their scope in the BSS/OSS domain, leaving less room to external system integrators. Telco's keep investing in BSS/OSS to improve customer experience, preventing churn, increasing revenue and reducing costs. In particular fulfilment and assurance get their full attention. The three main areas for improvement are: better service delivery, outsourcing / SaaS and (pre-paid post-paid) convergence. The focus is on end-to-end solutions driven by business values. With the cloud services as IaaS and SaaS more opportunities for complete outsourcing and off-shoring of BSS/OSS chains are created. Strategic partnerships with equipment and software vendors lead to combined proposals.

The last 10 years have shown a considerable consolidation of IT applications in the BSS and OSS domain. Companies like Oracle, IBM and Amdocs have acquired suppliers with packages for CRM, Billing and Service Management, integrating them into one suite. The required System Integration using EAI and SOA is part of the offered solution. This leads to offerings of larger IT blocks and

complete suites covering large parts of the BSS / OSS functionalities. The result is that now three different approaches for the BSS/OSS domain are followed:
1. Complete integrated suites, covering most of the required functionality
2. Separate applications in a standard fulfilment, assurance and billing architecture like TMF eTOM
3. Selected best of breed applications for separate BSS/OSS functions

Implementation and system integration of separate BSS/OSS IT applications will require a different approach than implementation of complete suites. Integration with legacy and migration will become the main activities. The system integration will take place at a higher level. Interaction with ERP systems like SAP and data warehouses of Oracle could be the core activities. Other tasks are setting of the suite parameters so that the service and business requirements are met. This means configuring the suite so that fulfilment, assurance and billing processes are customised and products, customer care and pricing plans are supported.
For System Integrators this means a different navigation and orientation; instead of focussing on individual packages, knowledge of complete solution suites becomes more important. For the Telecom sector this will require close cooperation and partnerships with the suppliers of these suites. The own development of bespoke solutions will be minimised. Even Adaptive Maintenance will be minimal as functional changes will be integrated in next releases of the suite suppliers.

IT maintenance can be reduced to parameterisation and operational technical management of the suite and platform. Partnership with the supplier of the suite should reveal the distribution of main tasks in managing the IT.
The result is more focus on Consultancy and Professional Services for the Telecom sector.

Figure 9 Navigation & Orientation

Important developments in the Telecom sector which have direct impact on IT support are:

> More video content, including more one-to-one or one-to-many video communications, more live video (e.g. concert and sport feeds), more niche content.

> Everything on demand, including all available video, music, gaming, corporate and personal content.

> Ubiquitous access, equating to seamless, affordable access on any access network, wherever the customer is located, including locations that are not part of the provider’s network

> Seamless, affordable access on any device, including PCs, TV displays, PDAs, mobile devices etc.

> More (and more varied) “over the top” services from the Cloud, in its widest sense. In time, the very term “over the top” is likely to lose its meaning as the distinction between Telco and third party services blur

> Targeted advertising and commerce as the commercial underpinning for more and more new services—replacing or augmenting direct payment for services

> Shortening shelf-life for certain categories of services, as the range of services widens and services come in and out of fashion

> More mash-ups and service blending, in the widest sense, blurring all existing service boundaries, including: launching one kind of service from another service, combining elements from one into another service

- Switching in session from one service to another, or adding service elements
- More services developed by end users, as well as adaptation of existing services to suit end user needs
- More machine to machine services sharing information among a network of devices.

Conclusions

Jacob’s staff has navigated us in the direction of future demand based on the emerging star technologies. Mobile broadband data growth has surpassed mobile voice. Convergence of Telecom Clouds is stimulated by demand but also by new technologies like LTE and FttH. TV and Video determine the required network capacity and access technology. Flat data rates will be limited so that the user will pay for what he needs. Outsourcing and Cloud computing become gradual more important for Telco’s. To participate in the value chain apps stores and SDPs play an important role. M2M provides new opportunities with fast growing sectors. In particular healthcare, transportation and payments are promising areas. New two-sided business models with shared risks and partnerships create new opportunities. The required IT support for BSS/OSS will change; larger integrated suites provide all required functionality, only customisation and connection with legacy will be needed. This all requires new navigation and orientation for Telco’s and IT suppliers. The reduced differences between IT and Telecom will lead to converging telecom clouds.

REFERENCES

TeleManagement Forum, App Stores: A Model for Future Service Provider Partnerships, USA 2010

TeleManagement Forum, Billing Evolution: The Lynchpin of Revenue Growth, USA 2010

UMTS Forum, Mobile Broadband Evolution: the roadmap from HSPA to LTE, 2009

UMTS Forum, Mobile TV advertising, 2009

Cave, J., van Oranje, C., Schindler, R., Shebani, A., Brutscher, PH-B., Robinson, N., Trends in connectivity technologies and their socio-economic impacts, July 2009

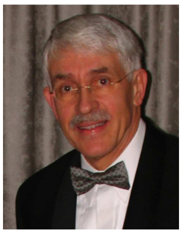
European Commission, Europe’s Digital Competitiveness Report, Vol 1, Brussels 2010

Nelson, E., van den Dam, R., IBM, Telco 2015 Five telling years, four future scenarios, 2010

Ovum Straighttalk, Telecoms 2020, Ovum Prespectives 2009

3GPP TR 23.879 3rd Generation Partnership Project Technical Specification Group Services and System Aspects, 2009

The author



Huib Ekkelenkamp graduated at Delft University of Technology in 1978 in the field of telecommunications. He joined KPN, with research on digital optical fibre transmission systems. He worked for the international consulting organisation of KPN in several countries. He spent many years in the Far East and worked in Indonesia in the area of telecom network planning. After his return to Europe he headed a KPN consulting team for telecom business customers.

He managed telecom consulting projects in Central and Eastern Europe and was involved in international acquisitions of KPN. In 2001 he became in KPN responsible for ICT business development. Currently he is in Atos Origin as Telecom Sector Manager with a team responsible for business and solutions development.

His main professional areas of interest are Fixed-Mobile Convergence, Next Generation Intelligent Networks, IP Multimedia Subsystems (IMS), Cloud Computing, Operational Support Systems and Service Delivery Environments.

huib.ekkelenkamp@atosorigin.com

APP ECOSYSTEMS AND THEIR RELEVANCE TO COMMUNICATION SERVICE PROVIDERS (CSPs)

Contributions to FITCE 2010 Proceedings

Bernd Wunderlich

IBM Deutschland GmbH, Hollerithstrasse 1, 81829 München, Germany

Bernd.Wunderlich@de.ibm.com

Keywords: Applications & Services, Emerging Business Models, Service Deployment/ Development, Application Stores, Apps, App Ecosystem, Wholesale Application Community (WAC)

Abstract: This paper analyses the market of App Ecosystems from the Communication Service Provider (CSP) point of view. It provides a market overview of App Stores, discusses possible future scenarios: Apple to keep its dominant position, Google to take major market share, and WAC to become market leader. It further discusses potential options for CSPs and concludes that the best option might be the WAC due to the gained market power for addressing the two sided business model and developer community which is addressed by the WAC. It concludes with the critical success factors for CSPs and WAC and highlights the capabilities WAC and CSPs need to have for playing in this market.

1 INTRODUCTION

Since Apple launched its Application Store or short "App Store" in 2008, the market for App downloads has grown to 4.1 bn USD in 2009 and is supposed to grow at 62% CAGR from 2009 to 2012 to a total of 17.5 bn USD generated by downloads for paid Apps and advertisement (Sharma 2010). Apple currently dominates the market with a market share of 93% in 2009. The downloads of Apps is supposed to grow with a 92% CAGR from 2009 to 2012 (Sharma 2010). Apple's success has seen a lot of companies including leading handset manufacturers, Communication Service Provider (CSPs), and independent internet companies trying to jump on the bandwagon.

The launch of the Wholesale Application Community (WAC 2010) on February 15, 2010 at Mobile World Congress (MWC) was the most important move of the CSPs moving into the direction of App Stores (TMF 2010). WAC has the goal to increase the overall market for mobile applications through the use of open standardized technologies, driving scaled deployment of those technologies and providing complimentary commercial models. WAC will merge OMTP BONDI and JIL as open Device APIs and will drive the development of a new DeviceAPI through W3C (WAC 2010).

This will allow developers to deploy a single application across multiple devices (through the use of standard technologies) and across multiple operators (without the need to negotiate with each of them) (WAC 2010). WAC claims that this approach will massively increase the number of applications which are available to users by simplifying their development, testing and deployment by developers.

A closer look at the App Stores reveals that there are currently 2 limiting factors for broader market growth of apps: 1.fragmentation of channels and platforms, 2. limited monetization models. The future of App Ecosystems will depend on various market choices and there is room for various possible scenarios. One possible scenario is that Apple keeps their market dominance. Apple has the unique advantages of a single most attractive device with the iPhone and iPad. On the other hand it is a closed ecosystem and with it limited to its iPhone and iPad market penetration and own distribution channels. Its key success factors seem to be the subsidization, of the App market in terms of iPhone sales push and attractive business models. Also Google is supposed to gain a significant market share with its Android system and App Store. Google could also become a single open ecosystem dominating the market similarly to how the internet has developed. On the other hand, WAC claims that they will create an open Ecosystem with open DeviceAPIs that would address a higher market than any of the other Ecosystem Players. Also WAC can leverage the existing contracts with device manufacturers and gain support for the APIs needed. With this, WAC would address potentially more than 3 Billion subscribers worldwide.

This paper addresses the following questions from the CSP point of view: What is the market for Apps? How can WAC and the CSPs be successful? What are the Critical Success Factors (CSFs)? What are the business models CSPs need to establish? What are the capabilities that WAC and the CSP need to develop?

2 Market Overview

As stated in the introduction, the market size for Apps has grown in less than 2 years to a market of 4.1 bn USD and is expected to grow to 17.5 bn in 2012 and to over 25bn by 2014 (Sharma 2010). This market can be divided in pure App sales and Advertisement. Advertising of Adds will play a much more important role in the future (Sharma 2010) and its share in the market is expected to grow from 12% in 2009 to 28% in 2012.
Of the mobile phone shipments in 2014, 37% will be Smartphones, the market share according to Frost & Sullivan will include 10% Apple, 12% Android, 42%, Symbian (mainly Nokia), and 17% RIM.
As mentioned, the market numbers provided reflects the revenue generated by direct sales and by advertising (Sharma 2010). Further revenue sources are not covered in these numbers but can be seen in various business models. Therefore, the indirect market is actually higher than the numbers mentioned above. For Apple, the increase of iPhone sales might still be more important than revenue generated by App sales.
The business models that are considered for marketing are: Paid Apps (Subscription, In-App or one time sale), Advertising (Impression-based, performance based, promotion), virtual goods, up-selling/ cross-selling other goods/ subsidization (e.g. mobile data plans), and hybrid (Sharma 2010).
Subscription and usage based apps and virtual currency are evolving business models (Cap Gemini Consulting 2010), whereas today mainly one-time sale, apps for marketing, and in-app advertisements are current business models.

Other business models that apply for marketing Apps include the Freemium business model (e.g limited functionality in free version).

Limited Monetization models for applications restrict development funding and demand for applications. Dominance of free to consumer retail models require the support of cross-subsidy business models to fund development similar to those developed for internet applications. The adaptation of most of these Business Models will be dependent on the Policies and Strategy of each individual CSP, which is likely to own the retail environments. The adaptation of a variety of business models is one Critical Success Factor (CSF) for CSPs to compete in this market.

It needs to be mentioned that not all of the revenue generated goes directly to the one who sells the Apps, i.e. the owner of the App store. Revenue sharing models with the developer apply which are as high as 80:20 for RIM and 70:30 for Apple and Google. This new business model can be regarded as the two-sided business model. The player in this market does not only compete for the App sale, but also for the developer who develops the applications. Apple e.g. has now over 225,000 Apps developed on their platform. This revenue sharing model attracted a lot of 3rd party developers (Sharma 2010), which is one of its key success factors.

As mentioned, there are various market players, also with new roles in the market. Device manufacturers introduced an Over The Top (OTT) business model, selling Apps directly or through third party app stores to the end user and making handsets more attractive through highly successful Apps such as Gaming.

3 Scenarios

Analysing the future by considering various scenarios is often done in strategic management (Thompson 2004). With the introduction of the WAC, there are 3 scenarios which are relevant for gaining a deeper understanding of the future market for App stores. These 3 can be described as:
- Apple dominant
- Google becomes market leader
-WAC gains market share and becomes market leader
These 3 scenarios will be described and discussed in the following.

3.1 Apple Scenario

With its closed Ecosystem, Apple continues to dominate the market similar to how it dominates music device and downloads. The iPhone and iPad are the leading edge and most sought after gadgets. Developers prefer the Apple app store as one single point of contact.
The lower cost of entry due to de-facto standard increases quantity and quality of applications. Apple has 225,786 Apps (http://148apps.biz/app-store-metrics/?mpage) and with it is the preferred App store for developers. It has also over 40,000 developers which publish Apps (http://mashable.com/2010/07/02/ios-android-developer-stats/). The various platforms cause confusion among developers. Developers therefore prefer Apple as a single Ecosystem. Apple is in a virtuous cycle, as they sell more iPhones with more Apps which brings more downloads and attracts more developers which write more applications.

Apple will significantly grow its market share of Smartphones. However, Apple still addresses currently only between 15% of the market for Smartphones (Gartner 2010). Other market numbers (Purdy 2009) show even that Apple's market share will continue to be around 10%. Even if this will increase to e.g. 20%, there are still 80% of the market to be addressed.

3.2 Google Scenario

Google is a single open ecosystem with its Android system dominating the market similarly to how the internet has developed. This increases revenue, applications and attracts developers and reduces overall costs. Android is adopted by various handset manufacturers. Android Apps are distributed via various channels. The free retail model remains the most common model, with the need for Advertisement or subsidization. The advertisement business model becomes more and more relevant and Google expands. However multiple versions of Android make it still difficult for developers. There are currently 85,000 Apps available and over 10,000 developers publish Apps. This number will with the increase of the market share of Android will increase. Due to Wikipedia (http://en.wikipedia.org/wiki/Android_%28operating_system%29) the Android market share is strongly increasing and already at 9% of the Smartphone market in February 2010. The forecast for 2014 shows that Android will be deployed on 12% of the handsets sold in 2014 (Purdy, 2010). For taking market leadership, this market share would have to increase significantly, which might be possible if e.g. more Handset manufacturers would join the Open Handset Alliance. Ovum (Ovum, 2010) predicts that Android will take 18% market share in the Smartphone market in 2014.

3.3 WAC Scenario

WAC is an open ecosystem which provides the "write once, deploy everywhere on any device" scenario for developers dominating the market enabling developers access to 3 Billion subscribers. This increases revenue and reduces overall costs (WAC, 2010). Addressing over 3 Billion Subscribers worldwide, more than any other environment, WAC brings down prices and makes itself attractive to developers more than other platform. With this gained attractiveness together with its global reach, WAC increases the overall market for mobile applications (WAC, 2010).
WAC will partner with leading device manufacturers including and serve a significant market share and with this will be able to address a major market share of the Smartphone market. The success of WAC depends highly on 3 factors: 1. which device manufacturers will support WAC and how big will be the addressable market, 2. to attract innovative developers, WAC will need to establish an attractive revenue sharing model. 3. the APIs for devices and network need to be enhanced and released quickly, 4. the governance of WAC needs to be handled with care, 5. WAC needs to act soon and become operational fast.
Given that each of these CSFs for WAC are key to manage in order to be successful, there is scepticism in the market that WAC will be successful. WAC will certainly need strong members and leadership in order to achieve its goals.

4 Challenges and Opportunities for CSPS

As mentioned two of the most limiting factors of the App market are the fragmentation of channels and platforms, and limited monetization models. The fragmentation of channels is addressed by the WAC by introducing a standard device API and enabling a write once, deploy on any device scenario. The second factor needs to be addressed by each CSP which will need to decide on the business model to market the Apps, but also the WAC needs to e.g. work together with Advertisers. In fact, it would be possible for WAC to address the gap that Advertisers want to launch campaigns for the entire market and not just for a single CSP. WAC could become the key interface of CSPs to the advertiser community and offer its inventory to advertisers in order to be able to launch e.g. multiple channel campaigns on a global scale. The huge amount of customer data is also seen by TMF (TMF 2010) as a key opportunity together with the already mentioned customer base and existing relationships. Another opportunity mentioned (TMF 2010) is the control over devices, due to the fact that most CSPs distribute and subsidize devices. This would enable WAC to put more pressure on device manufactures to support WAC and supports the WAC scenario.

The challenges that CSPs will phase in this market are mainly the time-to-market for establishing the App store in time (TMF, 2010). Also attracting 3rd party developers for writing applications for the variety of different platforms as device APIs such as JIL or BONDI are not yet mature enough and supported on many handsets. Also CSPs might not have the skills for handling partners such as developers and for testing and certifying Apps. One of the key challenges is also the support for a variety of business models and the unattractiveness of a single CSP to attract advertisers for launching single CSP campaigns.

5 Strategic Options for CSPS

The strategic options for CSPs that can be drawn from the discussions in the previous sections are:

- Do nothing
- Partner with third party app store
- Create own app store
- Partner with WAC

Each of these options has its advantages and disadvantages, but the option to Partner with WAC seems to be the choice with the highest success rate as it has the potential to attract a lot of developers which can write a lot of Apps. This can not be done by single CSPs or by a consortium of e.g. 3 CSPs such as JIL. It also solves the issue of skills for developing Apps, through bundling these into a single entity. If single CSPs create their own Ecosystem, they will have to invest a significant amount of financial and human resources. This together with the fact that it will be difficult for a single CSP to attract 3rd Party developers and to introduce the advertisement business model, which is becoming more and more important, leads to the conclusion that single CSP Ecosystems can not be profitable.

WAC seems to be the most likely scenario of choice for CSPs to compete in this market. However, CSPs also need to establish capabilities themselves, as WAC addresses the one side of the business model namely the developer ecosystem, but does not address marketing of the Apps. Therefore the CSPs need to also invest into own retail environments and introduce .

6 Capabilities needed by WAC and CSPS

This section assumes that the CSP partners with WAC and that WAC addresses the wholesale App part and handles developer relations, whereas the CSP handles the retail of Apps.

In this section, the capabilities needed to be successful in this market place, both on the WAC wholesale and a CSP retail model will be drawn from the discussion in the previous sections.

First of all WAC would have to build a WAC domain, or Wholesale Ecosystem, whereas the CSPs would each establish a retail domain with various different capabilities.

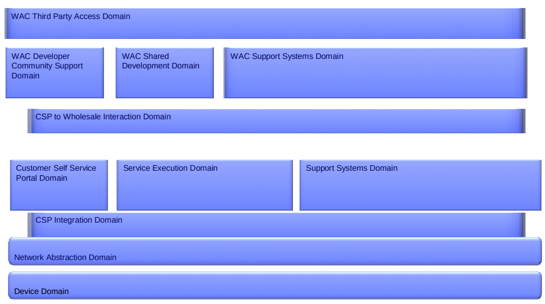


Figure 1: Domains of WAC and CSP

The WAC domain, which could be provided and hosted by a third party, using e.g. a cloud, consists of a Third Party Access domain and tools for provisioning content from 3rd parties and a Partner Portal providing a single interface to the WAC domain for developers and other partners such as advertisers.

The WAC developer community support domain is needed to provide collaborative tools such as Wikis and document repositories, but also reusable libraries for developing Apps to developers. Also the development environments need to be provided and managed. Open access for developers to network services of the CSP to create advanced Apps should be enabled as a key differentiator for the CSP.

The WAC shared development domain needs to provide tools for testing and developing Apps, such as device simulators, a Service Development Kit (SDK), the test environments. Not only the testing, but also the certification of Apps needs to be handled before they can be stored within a repository and made available to the CSPs which will handle the retailing.

The WAC Advertiser Support Domain needs to provide tools for Advertisers to launching campaigns, such as sales and campaign management tools.

Finally, the WAC support systems domain should provide Business Analytics tools for giving details on services to the developers in order to enhance services and focus on the most successful service offerings. The Information Infrastructure in terms of content registry & repository, user registry & repository also needs to be provided. It also needs to provide tools for commissioning of developers and other partners and for handling payment streams from CSPs. Also functions for security, for e.g. handling access control for developers, WAC employees, members, and partners need to be provided.

The CSP to Wholesale Interaction Domain will handle the payment streams from the WAC to the CSPs and from the CSPs to the WAC. Also the communication from customers and CSPs to the WAC, e.g. complaints or recommendations for enhancements, will be handled in this domain.

Each CSP needs to build also its own retail domain. The CSP needs to build capabilities for supporting the customer and handling customer requests and complains in the Customer Self Care domain.

It also needs to build capabilities to store, price, market, and download the Apps. An application front-end, App catalogue and App storage is needed. This will provide capabilities for the customer to search Apps, to recommend Apps to the customer and for handling the customer payment.

The Support Systems Domain provides capabilities for handling the communication and commissioning to the WAC and to other CSP systems such as billing. Capabilities for handling advertising campaigns such as ad insertion, campaign management, couponing, order fulfilment, ad portfolio targeting, usage analytics, and feedback and suggestions will need to be established.

The CSP Integration Domain will handle the integration of the systems and manage processes and data objects that are relevant to the CSP.
The Network Abstraction Domain will provide capabilities for providing network features and common enablers, such as location, or presence to the developer through a common network API. It is foreseen to use the GSMA OneAPI for providing these capabilities.

7 Conclusion

App stores are a very important component of the future mobile internet. CSPs need to play in this market if they want not simply become a bit pipe selling mobile data packages at market prices with limited control of their costs.
Currently 93% of the market share is attained by Apple, which have only 15% of the Smartphone market. Therefore, there is a market gap that CSPs and others can address.
CSPs have various options to play in this market. The level of success for these options remains currently unknown, but there is scepticism in the market that CSPs can be successful in comparison to Apple and Google. The level of success will depend highly on the ability to address the two-sided business model and attract developers and customers alike. WAC seems to the strategic option with the highest likelihood to be successful as it has the potential to attract developers solving a write once deploy on all platforms scenario, while addressing a market of 3 billion subscribers worldwide. CSPs should not see revenue of App sales as the only revenue source, but need to address a variety of business models.

ACKNOWLEDGEMENTS

I would like to thank my colleagues from IBM worldwide, who supported developing the IBM point of view for the Wholesale Application Community (WAC). Thanks for the many fruitful discussions on this highly interesting topic.

REFERENCES

Cap Gemini Consulting, 2010. *The Mobile Application Store Opportunity.* Presentation at the Tele Management World Forum, Nice.

Gartner, 2010. *Gartner Says Worldwide Mobile Phone Sales Grew 17 Per Cent in First Quarter 2010.* Gartner Press Release (http://www.gartner.com/it/page.jsp?id=1372013)

Leach, Adam, 2010. *Smartphones: the silver lining of the declining handset market.* Ovum Telecoms and Software News.

Purdy, J. Gerry, 2009. *2010 Outlook and Forecast – Mobile and Wireless Communication.* Frost & Sullivan Analyst Briefing.

Sharma, Chetan, 2010. *Sizing up the Global Mobile Apps Market*, Sharma Consulting.

TM Forum, 2010. *App Stores - A Model for Future Service Provider Partnerships*. TM Forum Insights Report.

Thompson, John L, 2004. *Strategic Management.* 4th Edition, Thomson Learning.

Wholesale Application Community, 2010. www.wholesaleapplicationcommunity.com

Zoller, Eden, 2010. *Navigating the Application Store Ecosystem.* Ovum Presentation at the Tele Management World Forum, Nice.

BUSINESS INNOVATION FOR A SUSTAINABLE INTERNET

Michel Defloor, Olivier Rits
Alcatel-Lucent, Copernicuslaan 50, Antwerp, Belgium
michel.defloor@alcatel-lucent.com, olivier.rits@alcatel-lucent.com

Keywords: Business model innovation.

Abstract: Current Internet business models are no longer sustainable while end consumers are unable to tap its full potential. To reach a sustainable business environment for all players across the value, companies should join forces and leverage each other's assets. There are 2 main options to enable such cooperation: dedicated partnerships, or open collaboration frameworks. To facilitate cooperation, the industry should structure an environment where business innovation is nurtured.

1 INTRODUCTION

For most people, the Internet is the same as the World Wide Web. This is reflected in the way one speaks about it: we surf the Internet, we search the Internet and we connect through the Internet. This situation however is changing.
Originally created for transport of data, the Internet enabled new applications such as e-mail, file transport, chat and the WWW. Today we witness three different waves in the evolution of the Internet:

2 three different waves in the evolution of the Internet

2.1 New services and applications

With the introduction of social networking technologies, the internet evolved in web2.0; and web3.0 (often referred to as the semantic Web) is knocking on the door. And there is more: IP networks have proven so efficient and successful that traditional telecom services now are also offered over IP: voice, TV, and video. And new services are fast coming our way: smart metering, e-health, e-government,…

2.2 Mobility: Internet on the move

With the introduction of the iPhone, we really entered the era of mobile Internet. Smartphones are a booming business and tablets are on the rise. It is hard to imagine a future where we won't be able to connect to the internet independent from where we are.

2.3 Machines: The Internet of Things

If you would think of the Internet as the brain of the world, the machine wave is all about providing the Internet hands and legs, by connecting it to sensors, actuators, and all kind of machines. This will also support an alternative definition of Web 3.0, being the Totally Integrated World.
Without most people realizing it, the Internet is more and more becoming part of our every-day lives: entertainment,

communication, information, collaboration, … Internet is getting omnipresent in a fast pace. A silent take-over.

3 Innovating the value chain

3.1 NEW KIDS ON THE BLOCK

This leads to a new business environment, where a new form of innovation will be required:
Where in the past, the internet was 'owned' by the ISPs, nowadays people associate it increasingly with household-names such as Google, Youtube and Facebook. Furthermore, a myriad of new application and content providers are jumping onto the bandwagon, offering their service over the WWW. Take the BBC for instance, that offers its programming directly via the internet. Even companies that were nowhere to be seen in online connectivity, such as car manufacturers, are now starting to embrace the web to enable totally new services.

This evolution puts pressure on the business models of most companies involved. Take for example the rise in data traffic needs, and the related required network investments by the telecom operators. Meanwhile the new kids on the block are able to bypass the telecom providers when it comes on revenue generation… The telco's need to invest, but don't see their revenue increasing! This type of issues triggered the net-neutrality-debate, but the concern is not limited to the telecom environment: take for example media-companies like music-publishers or newspapers, who see their revenue model challenged by online (legal and illegal) services. Every part of the value chain is today facing serious problems.

3.2 A new value chain

To reach a sustainable environment for all, a new value chain is required, and new business models need to be adopted. Only by joining forces and capabilities one will be able to unlock the full potential of the internet.
There are 2 main options to enable such cooperation: dedicated partnerships, where companies build alliances to cooperate on specific services (e.g. Verizon and Skype who cooperate to have a Skype-enabled Verizon mobile phone), or open collaboration frameworks where companies open up their assets to any potential partner in an ecosystem (e.g. open API services offered by GSMA). Both approaches have their merits, and what is the most appropriate option will depend on the relative marketposition of the companies involved, as well as their strategic direction.

3.2.1 Applications in Communications

This could lead to offerings of advanced communication services in multi-platform, multi-device environments. Today one has a unique opportunity to combine the best of telecommunications and the internet in a unified way, opening up advanced mobile communications to fixed-mobile convergence and web 2.0 integration. BT's Ribbit was an early example of this in the telecom operators' domain, but several other companies offer APIs in this domain.

3.2.2 Applications in media

Media-related applications will benefit from a coherent coordinated approach towards any screen. Legacy services are fragmented

across content types and devices — phone, PC and TV — limiting audiences and market potential. The industry can better monetize media assets, extend audience reach and attract new partners by automating the content life cycle and unifying content infrastructures.
The partnership between Youtube and Lagardere for the promotion and monetization of premium video content is exemplary in this field.

3.2.3 Applications in other domains

Many parties will have the opportunity of climbing the value chain by exposing assets & capabilities to partners, beyond current Web 2.0 collaboration:
One can to expose, in a managed and controlled way, selective capabilities, such as billing, context, security, quality, storage, subscriber preferences, etc.
Examples of such collaboration cannot only be found in the communications and media space, but also in the field of advertising, Machine-to-Machine communication (including smart metering) and eHealth applications. Take for example the cooperation between TomTom and various mobile operators (like Vodafone in the UK), to enable its HD traffic service.

3.3 Initiatives required at various levels

Business innovation is the main challenge today; as opposed to the often dominant focus on technical and service innovation. To facilitate this, initiatives should be taken at various levels:

- Companies should not only pro-actively include business innovation in their R&D effort, but embrace it as part of their DNA
- Governments can create a collaboration-friendly environment by acting on two levels: first they should provide for a regulatory environment to safeguard investments in open business innovation (including IPR), and secondly they can incentivize initiatives in this domain (e.g. via tax-incentives)
- Linked to government, education and academic research should shape a business mindset, on top of the topical learnings predominantly covered these days.
- Technology Federations can create platforms where the various players across the value chain can meet in a ‘safe' environment, beyond a single industry or geographic region. This includes standardization, but also initiatives like GSMA, FTTH Council, IAB, Greentouch, …

4 CONCLUSIONS

The internet will survive also without structured business innovation, but we believe we as an industry can offer the consumer a lot more value by cooperating across the value chain than we ever will be able to do individually.
There is a revolution hidden in collaboration. The telecom and internet industry has the choice to ignore it, or unleash it.

REFERENCES

Verizon Wireless and Skype join forces to create a global mobile calling community. http://about.skype.com/press/2010/02/verizon.html, Accessed June 20, 2010.

3rd Party Access Project - OneAPI https://gsma.securespsite.com/access/default.aspx, Accessed June 20, 2010

Lagardère Active and YouTube: strategic partnership for internet video broadcasting.
http://www.lagardere.com/press-room/press-releases/press-releases-363.html&idpress=4650, accessed June 20, 2010

INDUSTRIALIZED SERVICES ACCESS AND ONBOARDING

Enabling Applications While Managing Cost and Risk

Cindy A. Mills, Christophe Larcher

Alcatel-Lucent, 3 av. Octave Gréard, 75007 Paris, France

cindy.mills@alcatel-lucent.com, christophe.larcher@alcatel-lucent.com

Keywords: Onboarding optimization, Services Access and Onboarding, new business models, applications, application value chain, applications and content providers, monetizing, dynamic management, application enablement, industrialization, network provider, time-to-market, productivity

Abstract: Network providers are actively exploring new business models that will help them increase revenue and competitiveness through the delivery of new services and the enabling of large numbers of end user applications. Whether these models involve internal developers, a large number of third party developers, several aggregators, or only a limited selection of trusted partners or enterprise customers, all must meet stringent requirements for limiting capital expenses/operating expenses (CAPEX/OPEX), speeding time-to-market, and minimizing risk. Commercial viability requires the "industrialization" of Services Access and Onboarding (SA&O). This paper explains the why, what and how of industrialized SA&O.

1 EXECUTIVE SUMMARY

Network providers are actively exploring new business models to increase revenue and competitiveness by offering new services and enabling an explosion of end user applications. Whether the business model employs the use of internal developers, a large number of external developers, a few aggregators, or only a limited number of trusted partners or enterprise customers, all require employing new ways of keeping CAPEX, OPEX, time-to-market and risks low.

Enabling new applications requires evaluating the best way to expose value-added services and assimilate new applications and possibly new partners, into the network provider's environment in a controlled, secure, cost-effective and easy-to-use fashion. Alcatel-Lucent refers to this collective functionality as Services Access and Onboarding (SA&O). When SA&O is implemented effectively, benefits include:

- Reduced costs and time-to-market
- Improved quality and end user satisfaction
- Increased revenue through the leveraging and monetizing of existing network capabilities
- Improved productivity
- A strengthened application and content ecosystem
- Increased relevance of the network provider in the application value chain

Tapping into this new reality means that thousands of end user applications, and perhaps partners, will have to be supported at a rapid pace. Therefore, commercial viability requires the "industrialization" of SA&O.

At the same time, network providers have identified optimization of the onboarding process as one of the top two obstacles they face in growing this market. Depending on the business model, an industrialized approach will support characteristics such as, but not limited to:

- Large numbers of applications (apps) flowing into the system quickly and with very little administrative overhead
- Secure and controlled access to these new apps and the underlying network enablers
- Poorly performing apps automatically or semi-automatically being removed from the system
- Transparent, non-disruptive service and application updates
- Access to centralized data or automatically synchronized data (e.g., for customer relationship management, short code management, billing)

- Standardized and streamlined approaches for managing application and content providers (ACPs)
- Self-care portals for ACPs and end users
- Billing and value chain player payments occurring smoothly and promptly

Successful SA&O implementation requires a comprehensive lifecycle view of the business model requirements. Each element needs to be assessed according to impact on business goals and timeframe to complete, as well as complexity/risks and costs, keeping in mind the system, process, people and metrics implications. Industrialization needs to incorporate dynamic management, making billing support systems and operations support systems (BSS/OSS) “self-configuring”. From a strategic perspective, the business model determines where it is important to differentiate and where it is important to streamline and standardize.

The core of this paper provides an in-depth review of this issue. The introduction below addresses the why of industrialized SA&O, and the remainder of the paper focuses on the what and how.

1.1 Terminology Used In This Paper

Network Provider is the entity that delivers carrier network services in the wired and/or wireless market. Example services can include, but are not limited to, location, presence, voice call processing, delivery of videos to a mobile or IPTV device, subscriber information, and centralized billing.

Services access exposes network provider value-added services (e.g., messaging services, such as Multimedia Messaging Service (MMS) and Short Message Service (SMS), location information, subscriber data, on-demand bandwidth increases, centralized billing, customer care) using a well-defined, controlled, secure approach governed by pre-established policy and contractual agreements.

Onboarding is the assimilation of new partners, applications and service components into the network provider environment. For efficiency, it requires a set of formalized processes for applying the network provider’s governance policies to the initial approval, certification, assurance, provisioning and billing needs of the onboarded application or vendor.

Business model, when used in the context of this paper, addresses the business approach that includes aspects such as, but not limited to, revenue sources and charging options, policy and governance (including how rules are applied under varying circumstances, such as different branding and on/off-portal arrangements impacting what is expected), partner criteria and management requirements, service level agreement (SLA) considerations, retail/wholesale relationships and value chain, and services being offered.

Application and content providers (ACPs) represent a vast pool of entities that deliver value to end users by means of the communications services made available by network providers. As examples, ACPs include companies or individuals providing social networking applications, on-line gaming, news and sports updates, photo and video sharing sites, coupons and enterprise applications such as personal medical record management applications. Specific examples of popular ACPs include: Amazon®, SalesForce.com®, Skype, Google®/YouTube®, Hulu®, Apple® iTunes®, Flickr™, Twitter®, LinkedIn® and Facebook®. A network provider can function as an ACP through an internal service innovation and development organization and can enable its own services for end users.

Application Enablement is an industry vision and network approach that combines the trusted capabilities of network providers and the speed and innovation of the web to provide end users and enterprises what they demand: a richer and more trusted web and converged technology experience.

Application is a package (e.g., softphone, friend-finder) that delivers value to an end user, such as performing a particular task.

Service or service component is a capability (e.g., VoIP, location, presence, access to user preferences).

2 INTRODUCTION

The offering of new end user applications that operate across multiple network technologies (e.g., mobile and wireline) is happening at a phenomenal pace. Devices can include traditional phone sets, televisions and PCs, and they can also include game consoles, dog tags, electronic picture display frames, in-car Telematics (front seat, back seat) and medical devices. Much of this activity is being driven by the fast-paced evolution of web services, where openness, sharing, flexibility, convenience and innovation is thriving and impacting end user expectations.

Rapid change is underway. Today’s communications industry is seeing a flood of application stores being offered by mobile device manufacturers and the emergence of Internet players, such as Google and Skype, offering communication services across fixed and mobile devices.

Meanwhile, the use of mobile broadband is exploding. In Western Europe, for example, mobile broadband usage is predicted to have a cumulative average growth rate of 43 percent over the next five years (Yankee Group, September 2009). Smart-home technologies that will impact the use of broadband services, such as movie downloads, security videos, and utilities monitoring, will continue to drop in price and become more common.

At the same time, the competition in voice and data communications is causing prices and profit margins to drop. Therefore, it should come as no surprise that network providers around the world, such as Orange, Vodafone, AT&T, Verizon, and T-Mobile, have started or announced plans to offer new approaches to support the faster introduction of innovative, new services. The communications industry is just starting to recognize that new business models and their support systems are needed for network providers to stay viable.

Whether the new applications are developed internally or externally and regardless of the control requirements, the new business model approaches involve providing access to service components (e.g., resources) and the onboarding of new applications and perhaps new partners. The remainder of this paper will refer to this functionality as Services Access and Onboarding (SA&O). For additional reading related to the SA&O concept, including dimensions and inter-dependencies, see the documents listed in the references section of this paper.

When SA&O is implemented effectively, network providers recognize benefits, such as reduced costs and time-to-market, improved quality and end user satisfaction, increased revenues from leveraging and monetizing existing network capabilities, improved productivity, and increased relevance of the network provider in the application value chain. Therefore, SA&O can be considered both a defensive and offensive competitive play for network providers.

In 2009, 70 percent of network providers were onboarding less than five applications per month (See Figure 1). These results are based on an independent survey with over 100 responses, collected between June 24 and July 17 2009. Respondents were recruited from among more than 150 different network providers around the world. An effort was made to ensure adequate representation across all regions and across wireline, wireless, and cable network providers, with a deliberate focus on mobile providers. Multiple responses were allowed from a single provider, but the selection/screening process limited this occurrence to large network providers. Participants were pre-screened based on their knowledge of the third party application space. Example job descriptions included “head of portals”, “head of enablement strategy”, “content manager”, “application strategy manager”, “mobile Internet applications and products manager”.

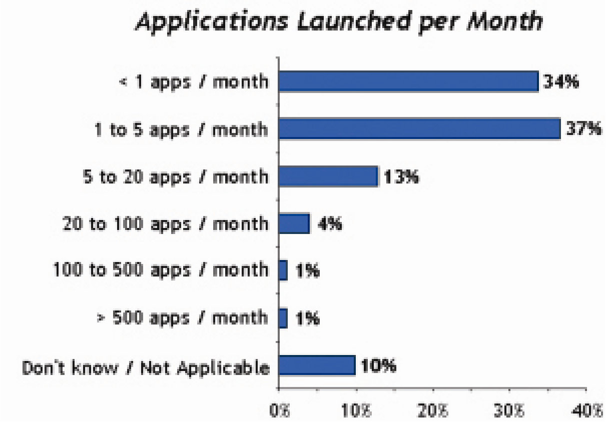


Figure 1: Number of applications launched by network providers.

Compare the results of this survey with Apple’s reports of receiving approximately 8,500 applications and updates per week. (Apple, 2009 and Fortune, 2009). And Facebook’s report of having 140 new applications added per day and 350,000 active applications (Facebook Developers Forum and Facebook Press Room Statistics).

The network providers who launch more than five applications per month tend towards one of two models:

- Operator-owned developer program (e.g., Orange Partner)
- Third party aggregator platform (e.g., Nokia Operator Partnership)

Furthermore, 60 percent of network providers rely on a white label approach for third party applications. A white label product or service is one produced by one company (the producer) that other companies (the marketers) rebrand to make it appear as if they made it.

This is still nascent territory for network providers, and everyone is learning as they go. Ways to make profitable revenue are shifting and expanding rapidly, (e.g., reflecting phenomena such as willingness to pay for low/no-cost digital products, such as $5 for an avatar or a package of seeds for a digital farm game and the growing free/inexpensive, on-demand mobile TV applications getting more popular than existing subscription-based broadcast mobile TV).

As a result, network providers are searching for business models that succeed, and, reflective of an experimental phase, some projects have been terminated abruptly, replaced, or have required large cost reductions for commercial viability. One of the key patterns emerging is that success requires a combination of operational flexibility and efficiency.

In recent surveys and interviews, network providers identified optimization of the onboarding process to allow developers to easily bring more applications to market as one of the top two obstacles they face in growing this market. The surveys showed that the number one obstacle was justifying the business case, unless the network provider had already chosen to move forward with SA&O, in which case the number one obstacle was identified as reducing network provider and device fragmentation to provide scale to developers. Interviews included 32 direct interviews performed by an independent company. Informal interviews by Alcatel-Lucent provided additional confirmation of the results.

Similarly, "integration into the network, operations and business systems," was identified as the number one pain point, with "standardization across devices and device O/Ss" a close second. Both of these results reinforce the need for an optimized, integrated and at least somewhat automated and standardized approach to SA&O.

For example, a Tier 1 network provider director recently mentioned that he compared his company's new application creation and onboarding process to having a dirt path through a forest. He explained that his company had established a narrow path, the first progress step, and could roll out a new application by taking a slow, rough journey. (He actually used the word "painful".) What he really wanted for applications enablement was a multi-lane, multi-ramp "super highway" (his term) with drive-through, automated toll booths. His immediate focus was strictly on internal development of new applications, but this discussion applies whether third party developers are used or if only internal developers create new applications. It's worth noting that a true super highway does require the use of a large number of developers to provide the innovation and volume required to continually provide new winning applications. However, it may be preferable to start with internal developers to build confidence in the approach.

What does an industrialized approach for the optimization of SA&O as a key to commercial viability look like?

Depending on the business model, example characteristics could include, but are not limited to:

- Large numbers of applications flowing into the system quickly and with very little administrative overhead. This implies:
    - Automated and continuous tracking and management of onboarding progress (including manual tasks, such as approvals)
    - Automated processes for provisioning and service assurance
    - Established options (e.g., for data access, revenue, billing) with associated controls and management built into the service models to support an "assembly line" approach
- Secure and controlled access to these new applications and the underlying network enablers
- Poorly performing applications being automatically or semi-automatically removed from the system, according to pre-established commercial and/or technical criteria
- Transparent, non-disruptive service and application updates
- Access to centralized data or automatically synchronized data (e.g., for customer relationship management, short code management, billing)
- Standardized and streamlined approaches (e.g., incorporating automation, reducing approvals, pre-established support tiers, incorporating communication-enabled business process management) for managing application and content providers (ACPs)
- Self-care portals for ACPs and end users
- Billing and value chain player payments occurring smoothly and promptly

Based on this overview of the why of industrialized SA&O, we can now focus on the what and how.

3 INDUSTRIALIZED SA&O SCOPE

There are multiple examples of network providers who have selected a few enablers, such as SMS,

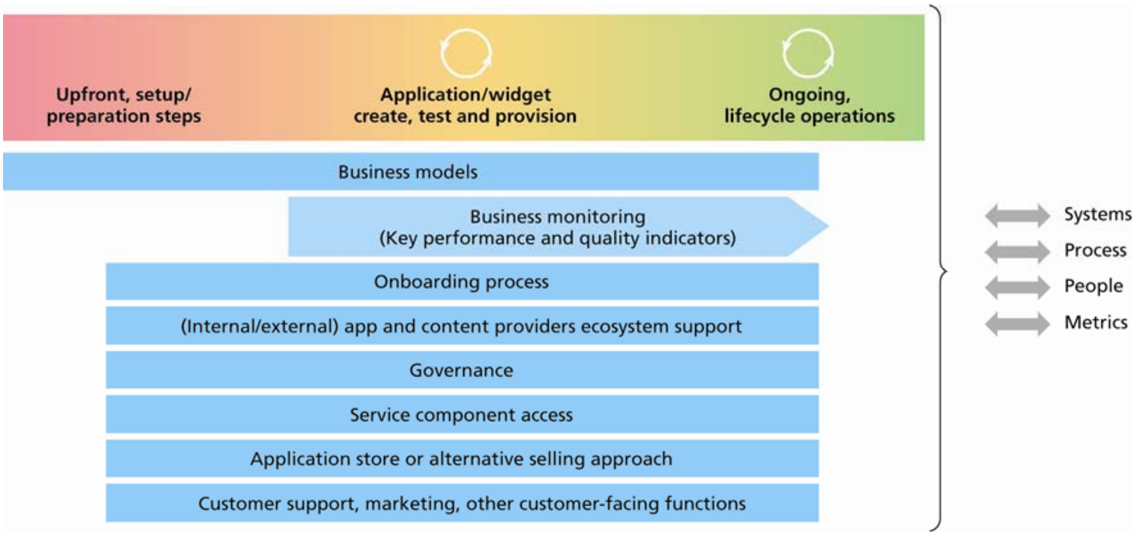


Figure 2: Dimensions of industrialized SA&O impact systems, process, people, metrics.

MMS, and WAP Push, and offered them to a broad external developer pool for application development. Real experiences show that when the planning and management only focused on the service exposure aspects, many network providers still took from five to 12 months to offer a new application on their system.

As Figure 2 indicates, industrialized SA&O involves multiple dimensions, all of which need to consider the systems (including data, security and legacy implications), process (both manual and automated), people (e.g., clients, employees, partners), and metrics (for performance and quality) requirements. Considering all these factors up-front supports high quality results, thereby meeting expectations of the key stakeholders, and lowers risks by reducing the chance for surprises.

While SA&O is primarily focused on the first two phases, the “On-going Lifecycle Operations” phase is directly impacted by SA&O. This is due to inputs from SA&O and for management requirements, and, therefore, cannot be completely ignored when establishing industrialized SA&O. In addition, continuous assessment and improvements are essential for all phases.

As Figure 2 illustrates, SA&O dimensions address:

- Identifying and implementing business models, including regulatory alignment, contractual agreements, revenue flow and settlements
- Establishing, collecting and monitoring key business performance and quality metrics
- Flexible service model-driven onboarding of new applications, services and perhaps partners
- Partner ecosystem support and management
- Governance of the design, onboarding and run-time environments
- Secure and policy-controlled access to service components
- Integrating with an app store or alternative selling approach system that includes application and service lifecycle management
- The full scope of customer relationship management, plus customer marketing

Today, most network providers employ a tightly-controlled and deliberate approach to exposing services and onboarding third party applications. An industrialized SA&O approach does not require elimination of that approach, although newly adopted business models may.

4 PLANNING AND IMPLEMENTATION

“Best-fit” starting steps vary by network provider and depend on considerations such as, but not limited to:

- Amount of manual operations the network provider can tolerate/afford
- Extent of systems, back-office tools and processes additions, modifications or integration
- Business model and partner requirements
- Budget or other restrictions

To identify low-hanging fruit (i.e., relatively low cost, fast and positive impacting) or to work items with the highest impact, there must be clarity (and ideally alignment) on business goals and business models. Once that is established, the first step involves taking a closer look at the SA&O dimensions mentioned in the previous section.

Particularly important is the introduction of dynamic management. Along with standard industrialization approaches, such as BPM and automation, dynamic management introduces the concept of “self-configuring" BSS/OSS systems. This means automating configuration, activation, billing and assurance, as well as incorporating dynamic service governance for each new service, application or partner. This reflects a framework of pre-established service models with operational profiles that apply industry best practices and tools, such as Service Oriented Architecture (SOA), Information Technology Infrastructure Library (ITIL) and TeleManagement Forum (TMF). The result is a fast, secure and scalable solution that enables the deployment of new services without exponential OPEX growth.

Industrialization with high quality of experience (QoE) goes beyond system requirements and addresses the people, processes and metrics across the full lifecycle, from service introduction, customer order, service guarantee and troubleshooting to customer care, billing, and revenue sharing with third parties, and service retirement. This means eliminating inefficient silos and the error-prone manual and one-off, customized approaches.

This high-level, simplified view might be interpreted to imply that industrialized SA&O is simple for a network provider to implement internally. While it’s true that numerous, commercial, off-the-shelf products exist for functional areas of a SA&O solution, and that a network provider’s internal staff is highly experienced and the most knowledgeable about current infrastructure and processes, it’s also true that applying direct experience and lessons learned in this space will reduce risk and increase the likelihood of successfully meeting goals.

For example, understanding developer expectations allows the implementation of an infrastructure that will encourage, rather than discourage, broad acceptance from the developer community. Also, numerous companies have introduced the use of SOA approaches only to realize one or two years later that expectations were not met due to forgetting key SOA elements (e.g., governance components) from the start.

Another example consideration is aligning the network, IT and operations processes and staff to avoid missing key requirements, migrations and operational efficiencies. Planning a phased approach for system configurations, incorporation of metrics and operational changes requires establishing a team with end-to-end skills, pre-established partnerships and insights into lessons learned from previous, related experiences.

The reality is that SA&O solutions are operationally complex and, therefore, need to be industrialized because they operate across technologies and organizational boundaries. They are likely to be:

- Partially traditional telecommunications infrastructure
- Partially IT servers in data centres
- Partially at partner premises
- Partially in a customer home network
- Location, device and access independent
- Supportive of varied SLAs and business models

This type of situation requires a combination of network, IT and operations expertise and the incorporation of corresponding industry best practices.

5 CONCLUSION

Success with new business models associated with enabling large numbers of new applications requires understanding the implications of the business models and an industrialized approach. Focusing only on exposing service components (i.e., resources/enablers) has been shown to produce disappointing results. What’s required is a comprehensive, full-lifecycle view of the business model requirements. This allows the creation of a complete end vision for the technical and operational structures. Once the end goal is well understood, then the network provider can determine the best approaches, starting points and the phased program to meet the goals.

ACKNOWLEDGEMENTS

The content of this document reflects collective thinking and experiences of the Alcatel-Lucent Multivendor SDP and SOA competency center and BSS/OSS competency center team members.

REFERENCES

Yankee Group, September 2009. *EMEA Mobile Forecast.*

Apple, August 2009. *Apple Answers the FCC's Questions*, http://www.apple.com/hotnews/apple-answers-fcc-questions

Philip Elmer-DeWitt, August 2009. *40 staffers. 2 reviews. 8,500 iPhone apps per week*. Fortune. http://brainstormtech.blogs.fortune.cnn.com/2009/08/21/40-staffers-2-reviews-8500-iphone-apps-per-week

Facebook, *Facebook Developers Forum*. http://forum.developers.facebook.com/viewtopic.php?id=15866

Facebook, November 2007. *Facebook Press Room Statistics*. http://www.facebook.com/press/info.php?statistics

Wikipedia, http://en.wikipedia.org/wiki/White-label_product

Cindy Mills, July 2009. *Services Access and Onboarding: The New Game for Network Providers.* Alcatel-Lucent.

Patricia Hargil, Chris Kapuscinski and Ann Marie Vega, March 2010. *Crafting the Network Provider's Business Case for Application and Content Provider Collaboration - challenges, risks and key success factors.* Alcatel-Lucent.

Patricia Hargil and Ann Marie Vega, February 2010. *App Stores Everywhere: How Innovative Network Providers Are Standing Out From the Crowd.* Alcatel-Lucent.

Graham Finnie, May 2009. *Working with Third Party Services: An Action Plan for Network Providers.* Alcatel-Lucent.

Alcatel-Lucent, 2009. *Application Enablement: Delivering a Trusted Web Experience to Increase Revenue.* Alcatel-Lucent.

HOLISTIC APPROACH FOR IMPROVING THE FTTH BUSINESS CASE

Koen Casier, Sofie Verbrugge, Bart Lannoo, Jan Van Ooteghem, Piet Demeester
Dept. of Information Technology (INTEC), Ghent University/IBBT, Gaston Crommenberglaan 8, 9050 Gent, Belgium
{Koen.casier, Sofie Verbrugge, Bart Lannoo, Jan Van Ooteghem, Piet Demeester}@intec.ugent.be

Keywords: Techno-Economic, FTTH, geomarketing, synergetic installation, operational expenditures

Abstract: For most European operators, there is more value in upgrading their existing infrastructure than installing a new FTTH network. One approach to improve the FTTH business case could consist of using a holistic approach, where different aspects are optimized together.

First the operator needs to know which areas in a region to rollout first, to postpone or skip. The operator will select the best set of customers to connect in order to maximize its business case. Next, the operator should look for approaches to reduce the costs of an FTTH rollout. The cost for installing the fibres in the network is dominating and the largest reduction here can be achieved by finding synergies with other infrastructure owners. Finally the operational expenditures are also an obstacle for many operators. They have a good understanding of operations for their current infrastructure, while a new fibre based infrastructure will clearly bring uncertainty here.

In this paper, we show how tackling the business case on those three fronts – strategic geo-marketing, synergetic installation and detailed operational modelling –greatly improves the viability of the business case of FTTH. It could also lead to an earlier FTTH deployment and higher coverage.

1 IMPROVING THE FTTH BUSINESS CASE

In many European countries, there is already a good existing copper or coaxial telecom access network in place [1]. Replacing this network by a fibre network up to the customers' homes, involves tremendous infrastructure works. For most European operators, there is more value in recycling and upgrading their existing infrastructure, regardless of the higher equipment and operational expenditures, in order not to incur the large trenching overhead. Still, as soon as one operator deploys FTTH, all other operators will most probably follow and aggressively rollout FTTH in the same regions in order not to lose a foothold there [2]. Each operator must be prepared for such situation and have a solid future proof strategy for the rollout. To improve the business case of a FTTH rollout, we believe that there is a need for a holistic approach where different aspects are optimized in a common way instead of considering them separately.

The operator needs to know which areas in a region to rollout first and which areas to postpone or skip. The customer is of vital importance in the outcome of the business case, and the operator will have to select the best set of customers to connect in order to maximize its business case. Working at this level requires a huge amount of information and calculation, and this quickly becomes prohibitive. Building such a geo-marketing strategy requires intelligent clustering approaches aimed at reducing the complexity while not discarding too much detail. This process should be split in three consecutive steps: (1) aggregating all input information and extracting a logical classification of the customers in logically separated types, (2) clustering groups of customers according to their profile and the trenching distance required and (3) extract the best rollout strategy based on those groupings.

The cost for installing the fibres in the network, also referred to as the outside plant, will be dominating (e.g. in a fully buried installation this can amount up to 70% of the overall costs). The business case can be substantially improved by lowering this installation cost. The largest reduction in cost can be achieved by finding synergies with other infrastructure owners for the installation of the network. When the installation in the trench can be completely split between two operators, the cost for each operator will drop to almost half. Three important questions arise in this context: (1) how can the different operators be encouraged to cooperate for installations, (2) how can the installations of different infrastructure owners be optimally synchronized and managed, and (3) how should the costs be split amongst the different operators in order to fairly reflect the joint and dedicated installation part of each infrastructure.

Finally, next to the trenching cost also the operational expenditures (OpEx) will be an important obstacle for the operator. The current copper or coaxial infrastructure has been in use for several decades. The operator has a good understanding of the operational processes, expenditures and optimizations in its network. With the advent of a new fibre based infrastructure, this brings additional risks into the corporation. Clearly considering the distributed character of the outside plant, this cannot be neglected. Even more, when different infrastructures are combined in trenching, ducting, or even up to installation, the new operational processes are hardly known.

Tackling the business case on those three fronts – strategic geo-marketing, synergetic installation and detailed operational modelling – simultaneously would greatly benefit the viability of the business case of FTTH and could lead to an earlier FTTH deployment with a higher coverage. In the following sections, the opportunities on each of the three considered fronts are discussed in more detail.

2 FOCUS ON THE BEST CUSTOMERS

Any business case will start from the customers. Who are they and what are they willing to pay for the products or services? An overall view on the customer base could be sufficient for a low risk deployment. For a project involving huge upfront investments, such as an FTTH deployment, the profile of the customers should contain as much detail as one can get. The vast amount of information on each potential customer will be too much to understand and work with. The operator will typically try to identify profiles of customers, e.g. young-telecom-minded, early-adopter, IT-professional, video-enthusiast, etc. For each of those profiles, the operator can build a detailed marketing strategy, adoption model, etc. and unify them in a dedicated business model. However, as FTTH is a fixed architecture, it involves large infrastructure works for providing all customers with a fibre connection to the central office. This part of the access network contains all fibres, cables, and ducts, is called the outside plant. The structure of the outside plant renders it infeasible to connect on a per customer basis. The large costs for connecting the customers will call for a further clustering exercise in which the best groups of customers are selected according to their profile (average for the group) and the expected installation costs.

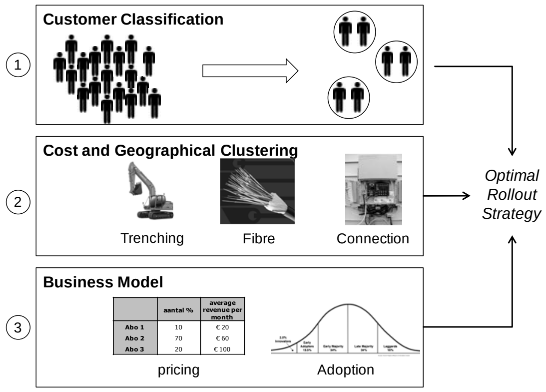


Figure 1: The three geomarketing steps to come to an optimal rollout

Figure 1 shows the three steps in constructing the best cherry picking strategy for the operator: (1) gathering all information and classifying all inhabitants according to their profile, (2) grouping customers in selected areas according to their profile and expected trenching length and .(3) making up the full business model given the customer groups and priorities.

As a basis for the cherry-picking, the operator has to find all information available for each potential FTTH customer (both inhabitants and companies). Considering an FTTH deployment the following five classes of information will be very important:

User related information

1. Demographic information: type of inhabitants, family situation, type of dwelling, education etc.
2. Economic information: average income of the inhabitants, type of business, turnover of the companies in the area, etc.
3. Marketing information: existing customers, customer base for any competitor, average revenue per user (ARPU), value adding services over the infrastructure, etc.

Infrastructure related information

4. Geographic information: is there infrastructure available to be reused, will the cabling be deployed on the sidewalks or at the edge of the street, the type of soil and existence of barriers for crossing (e.g. major roads, rivers, etc.

5. Other sources: own infrastructure such as buildings, street cabinets, existing ducts, etc.

This information must be aggregated within one database and all missing information must be clearly indicated. An operator will typically have to contact a third party for retrieving this information from the different sources and even from its own data-warehouse. The existence of such a data-integrator and the amount of detail in which it can offer data will be one main determining factor for the pace at which the operator advances in planning the FTTH deployment strategy. The information in this database will without doubt contain too much detail to work with. Especially for reducing the complexity of the calculations, it will pay off to aggregate the existing data. As the outcome of the business case is most dependent on the customer revenues and the overall costs, these will also be the main directions to look for in step 1 and 2 in Figure 1.

In the first step the customers are grouped according to a limited set of profiles. To this goal, different techniques from data-mining are used. It is for instance useful to use statistical methods for finding highly correlated values, leading to a limited set of customer types (profiles).

In the second step, the essential and specific background of the FTTH business case is brought into the geomarketing calculations in order to really grasp all possible cost reductions. This is typically tackled using some kind of geographical clustering approach. Clearly customers will be grouped according to geographical distance and their assumed ARPU as found in their profile. As such we will most probably find different smaller closed areas in the region in which all customers have (or lack) more or less the same drive towards FTTH.

Finally, at the end of the first two steps, the operator ends up with a data-set containing for each customer its geographical group and customer profile it belongs to. This information will form the basis for calculating the cost for the deployment of FTTH in each part of the region. By deploying each group at the right time taking into account the full business case will provide the optimal rollout strategy for the considered region

In [3] we investigated the potential of geomarketing for improving the FTTH business case. We found a huge improvement for a geomarketing FTTH business case, in comparison to the original business case which was developed in [4]. The results indicate that geomarketing could increase the final outcome of the FTTH rollout with more than 20%. Additionally it shortens the payback period by two years and reduces the initial investments up to 20%.

It is important to note here that the tools and the manner, in which the integration from information, through customer profiling up to deployment strategy is handled, will be of high importance. With a flexible and extensible implementation in place, the operator can perform the study again at a later stage without much effort, reflecting changes in customer information, equipment pricing, installation costs, etc. Iterating over this approach with more detailed calculation approaches, both in clustering as in estimating the installation length and cost, allows for reaching a highly reliable strategy for the rollout of an FTTH network. Additional opportunities for savings in trenching of the outside plant are discussed in detail in the following section.

3 DEPLOY IN SYNERGY

As mentioned, the costs for rolling out FTTH will be dominated by the costs for installing the outside plant. Especially in the European context, where trenching is often required by law, the digging works are prohibitive for all but the very high density cities. Any possibility for reducing this cost can greatly improve the business case for the operator. Technological advances will lower the costs for the equipment to install in the network and will without doubt deliver new and enhanced installation techniques such as micro-trenching. Much more savings are possible by looking for synergies when installing the infrastructure, as shown Figure 2. A joint installation or network sharing between different operators will allow all operators to reach the customers at a joint cost of only a single (albeit possibly slightly more costly) installation. By sharing this cost, the dedicated cost per operator might decrease significantly. Joining forces with other infrastructure providers, e.g. gas or electricity (see also Figure 2), leads to comparable savings. It’s worth noting that the synergy can stretch up to the customer connection, in which the customer is connected to all infrastructures in only one intervention. This could again save a lot, but might require additional administration, aligned operational processes and more trained technicians.

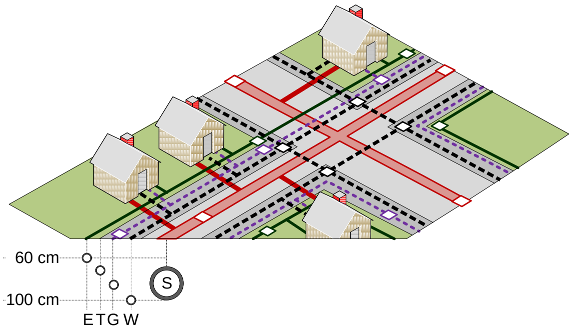


Figure 2: The short proximity of underground infrastructures (Electricity (E), Telecom (T), Gas (G), Water (W) and Sewage (S)) allows important cost savings through synergies.

Not only for operators but also for public authorities, it pays to facilitate the cooperation between the different infrastructure owners, as this will reduce the final costs charged to the customers as well as reduce the amount of road works in the area. Clearly many different parties will be involved and can gain by joining the installation partly or completely in a given region. It is essential to draw a complete value network involving all of them in order to take the right decisions. A so called multi-actor analysis will clearly show where the different actors are performing comparative roles over the different infrastructures, and where the main differences are found. The analysis helps in identifying the instances in which a joint installation of FTTH with other parties might be possible and provide cost savings for all parties. Here lies an opportunity for public actors to improve the cooperation and interaction between the different players.

Once all actors have been identified, it will also become clear at which level the cooperation can take place. Figure 3 gives an overview of the different levels at which two or more infrastructure owners can jointly deploy their network.

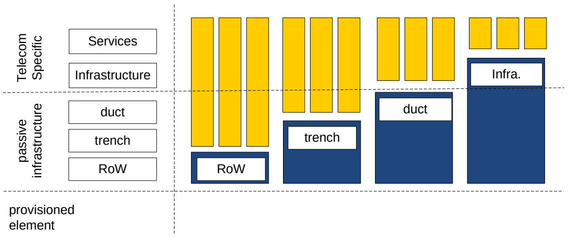


Figure 3: There are cooperation and unbundling opportunities at different levels.

Cooperation at the right of way level is a fairly straightforward step in which the public authorities can be a moderator or an actual party involved. Lowering the requirements for acquiring the right of way for joint deployments for instance can force the infrastructure owners to cooperate. At this level also a broad platform of actors can open up a broader right of way for instance next to railway infrastructure or highways, in existing ducts or on the facades/poles.

Cooperation with joint trenching will most probably lead to the highest cost reduction per party. To put things simple, installation of two infrastructures in the same trench will reduce the dedicated trenching cost to half. This is of course an overly simplified example as a joint installation will lead to additional management and administration. Additionally, the planning process should also take this into account as different infrastructures will have diverse requirements on installation (e.g. maximum number of customers on a line, legally enforced minimal safety installation depth and distances between infrastructures, maximum length between customer and central office, number of control points per km, etc.). In [6] we performed an initial quantification on the costs of a joint installation at this level. The results of this study showed that more than 20% and up to 55% installation cost savings were possible for a joint installation in a dense urban situation and 10% up to 45% in a more rural area. Most savings were possible in case all infrastructure owners – gas, water, electricity and telecom – would join the installation. In case this is not possible, the second best synergy is found between telecom and electricity when considering an FTTH network. This joint trenching is currently only realistic in Greenfield situations or large municipal works. This study only looks at the expenditures in trenching and installation, while cooperation at this level will also have an impact on the operational processes, as it might imply a different topology, control locations, lower distance between cabling, etc. This impact will be discussed in more detail in the following section.

Cooperation at duct level involves the joint installation of a duct topology in which different infrastructure owners can install their dedicated equipment. Synergy at this level resolves some of the difficulties as mentioned in the trenching level cooperation (easier administration, cost allocation, etc.). Still a lot of technical issues need to be resolved for such a far-reaching cooperation. Here especially the vendors of passive equipment and cabling, as used in the outside plant, can take actions to facilitate synergies at this level. Cooperation at the ducting level will without any doubt have a considerable impact on the operational processes. Currently often only telecom infrastructure is using ducts and those ducts are not reusable for other infrastructure owners.

At a higher level, the cooperation is typically restricted to telecom operators sharing the fibre. Regulatory instances often act at this level and can force the operator to open up the network at a predefined level (fibre, wavelength, bitstream, etc.) for a cost-based determined tariff.

Deploying FTTH in synergy with other operators or infrastructure owners holds the promise of reducing the cost of the outside plant - the dominating cost (up to 70%) - considerably. Still there are several obstructions to be tackled, both of technical, operational and administrative nature. Public authorities can play a very important role here in identifying the actors involved and giving (e.g. legislative) the right incentives to cooperating actors. In addition, it is important for all actors to agree on which level the joint installation should take place and how the costs will be divided amongst all. Synergies will benefit substantially when the trenching or ducting level are shared.

4 OPERATE THE JOINT INFRASTRUCTURE

Installing the outside plant in cooperation with other operators and infrastructure owners will definitely reduce the costs for each actor. On the other hand, such joint installation will undoubtedly lead to important questions considering the operations of the network. In case of an FTTH network deployment, the operator will replace large parts of the access network with a new fibre based network. It has no long-running experience with respect to maintaining and replacing the fibre based access network and equipment. Operational expenditures can sum up to 50% of the total costs; still they are often modelled in little to no detail. Without any doubt this uncertain situation poses additional risks.

In order to accurately model the operational costs, two aspects are essential: the flow of activities in the considered processes needs to be detailed and the required input data has to be estimated. The aim of this modelling is to construct a large overview of all processes taking place in the network, and to make an accurate presentation of the different steps taking place in each of those processes. At the same time, the influence of a joint installation on the operations has to be kept in mind.

Key is understanding that, although FTTH is an entirely new network, a lot of information is readily available from the existing network. Many of the steps taking place in the processes for an existing network infrastructure will also be required in the corresponding processes in an FTTH network (or in a joint network infrastructure). Although there is a difference between copper welding or coaxial repair and fibre splicing, many other activities in the failure repair process will still be very similar for FTTH, e.g. the time consuming activity for the repair teams of getting to and from the location of the failure. It is advisable to use a modelling language (or graphics) that is intuitive to the different people involved in the process (e.g. technicians, experts). Typically flowchart based approaches are used for this case. They are well standardized, fit intuitively with existing information sources and are easy to understand. Plenty of tools are available for modelling and drawing operational processes.

Figure 4 gives an example for the process model used in [7]. This process model shows the actions to be taken when an underground infrastructure is damaged. We used this model as a starting point and made a quantitative cost comparison between a separate repair process and a joint repair process for all infrastructures - telecom, electricity, gas and water. The results showed that, when infrastructures are not close to each other and consequently do not fail together often, this joint process would increase the operational expenditures for the repair. On the other hand, when the infrastructures are close to each other and have a larger possibility for failing together, the joint process can become more cost-effective. In the situation where the infrastructures always fail together, a cost saving of the repair process up to 40% is possible.

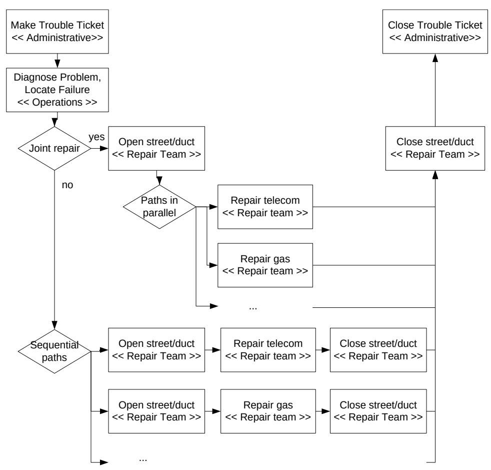


Figure 4: flowchart model for the repair process with an indication of the difference between joint and separate repair actions.

It is essential for an operator to model all operational processes in detail. Once modelled, the costs for executing these processes can be estimated fairly straightforward. Especially considering synergies and the repair of network failures, a lot of questions remain unanswered in the process description. In case of a joint installation, and especially in case of shared ducts, special attention should be given to the flow of control between the different infrastructure owners. Which infrastructure has automated fault detection and at what accuracy, how will every actor involved (also public actors) get informed of the problem, who will perform repair actions first, who will dig up to the cables and how will this cost be allocated to the others using the same well, etc. More detailed modelling of all operational processes will also enable the operator to compare more trustworthy different alternative installation types or operations, to find bottlenecks and scheduling problems. As such he can keep a good eye on the current OpEx and control and optimize future OpEx.

5 A CALL FOR ACTION

Clearly a holistic approach – where different aspects are optimized in a common way instead of considering them separately – can substantially improve the viability of the business case for an FTTH deployment. As a consequence it could help to speed up the rollout of FTTH in Europe and lead to a higher coverage at the same time. It requires actions of the different actors – operators, infrastructure owners and public instances - involved.

The first focus is on the overall business case for which the customers (ARPU, adoption, etc.) play the most important role. Public instances can enhance the view of the operators by facilitating information gathering on (potential) customers. Dedicated data integrators would also be very valuable in filling a gap here. The operator can use this information in combination with techniques from data-mining and geomarketing to find a viable long-term FTTH deployment strategy. Initial research showed a great potential improvement here, with an increase of the outcome up to 20% in comparison to the original business case.

The second focus is on reducing the costs, especially the dominating costs for the installation of the outside plant. Here, synergies with the other infrastructure owners and public authorities can help to significantly reduce the installation costs. The public authorities have an important role here. They can promote joint installations by facilitating the right of way; they can moderate the synergies between different actors on other levels and can finally also put the right incentives (e.g. legislative) in place to force the different actors to cooperate. Initial research showed a potential cost savings in the range of 10%-50% when different infrastructure owners would perform joint installations of their equipment. The full benefit can be obtained when all infrastructure owners (gas, electricity, tap water and telecom) join installation.

Finally it is crucial to reduce the risks of the new FTTH network infrastructure to a minimum. In this context especially the highly uncertain operational expenditures are very important. Operators need to construct clear models for the operational processes taking place in an FTTH or joint network. They can readily gather a lot of information on this from their existing installation base. Finally also the vendors should provide clear operational requirements for their technical solutions. Initial research showed here that savings are not always possible, but could in some specific cases amount up to 40% of the original expenditures for these operational processes.

Clearly while all three approaches, especially in combination, promise important cost reductions, there is still a lot of research to be tackled on each of them. This does not only require additional theoretical research, but will also require improvements in logistics, administration and management.

ACKNOWLEDGEMENTS

The author would like to thank all students and colleagues that were involved in testing the claims and recommendations mentioned in this paper.
The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 249025 (ICT-OASE) and was also carried out in the framework of the COST ISO605 Econ@Tel.

REFERENCES

[1] Point topic, World Broadband Short Report, http://point-topic.com/content/dslanalysis/World Broadband Short Report Q1 2010.pdf, 2010.

[2] K. Casier et al, *Game-Theoretic Optimization of a Fiber-to-the-Home Municipality Network Rollout*, Journal of Optical Communications and Networking, Vol. 1 Issue 1, pp.30-42 (2009)

[3] Manon Sikkema, "*Geomarketing technieken bij het uitrollen van nieuwe netwerkinfrastructuur*", Ghent University Master Thesis, 2010. <Dutch>

[4] Bart Wouters, "*Invloed van stadsnetwerken op de breedbandmarkt gemodelleerd met behulp van speltheorie*", Ghent University Master Thesis, 2007. <Dutch>

[5] B. Lannoo, et al *Economic Benefits of a Community Driven Fiber to the Home Rollout*, Broadnets 2008, London, Sept UK, 8-11, 2008

[6] Lukas Deschuttere, "*Multi-actor analyse voor een synergetische samenwerking van verschillende nutsnetwerken*", Ghent University Master Thesis, 2010. <Dutch>

[7] Mathias Gheeraert, "*Het optimaal op elkaar afstemmen van de operationele processen binnen verschillende nutsnetwerken*", Ghent University Master Thesis, 2010. <Dutch>

PROSPECTS ON FTTH/EP2P OPEN ACCESS MODELS

Adrián González, Arturo Vergara, Antolín Moral, Jorge Pérez
Signal, Systems and Radiocommunications Department, Universidad Politécnica de Madrid (UPM), Av. Complutense nº30 C-431, 28040, Madrid, Spain
adrian.gonzalez.alvarez@gmail.com, {avergara, amoral, jperez}@gtic.ssr.upm.es

Keywords: Ethernet point to point, fiber to the home (FTTH), next generation access networks (NGA), open access models, techno-economic analysis.

Abstract: In this article we will analyse, through a techno-economic perspective, the viability of the application of open access models to the rollout of FTTH networks based on point-to-point architectures, and compare it with traditional models. To that end, we have developed a specific model in the framework of the COSTA Project, which calculates the deployment costs related to the rollout of these networks at the different access levels in different geographic scenarios.

1 INTRODUCTION

The proliferation of Ethernet LAN's and the improvements in core networks are creating a bottleneck in the access network that telcos will have to address to remain competitive in the mid term. Next Generation Access Networks (NGA's) appear as the logical evolution to the current access technologies but few operators can face the huge investment effort required, especially nowadays.

European perspectives for a fast and large-scale NGA deployment in general and for the fiber-to-the-home (FTTH) networks in particular, are not too optimistic compared to worldwide leaders as Japan, US or Hong Kong (IDATE, 2010), due to several uncertainties that limit or hold these rollouts (Noam, 2010). The main inhibiting elements are the lack of clear demand for the new services and capacities (factor that sees itself amplified by the current financial crisis), the difficulties for the operators to obtain an adequate ROI, as well as the regulatory risk. Thus, the huge capital investment needed for the rollout of these new networks does not seem to be aligned with operators' profitability objectives for the short and the mid-term.

To deal with these uncertainties and in order to improve the return of investment, some operators (such as KPN) are introducing new models in which the deployed networks are voluntarily offered to alternative agents. Under these models, usually called open access models, operators are trying to increase the number of services offered through the network directly or indirectly, services that can substantially improve the business case of the FTTH deployments compared to vertical approaches. This improvement of the business cases could lead to a faster deployment of these new networks and, consequently, to a higher competition at the service level, increasing the network value for end users.

The open access model can be better understood through Ad Scheepbower's (CEO of KPN) words: *"In hindsight, KPN made a mistake back in 1996. We were not too enthusiastic to be forced to allow competitors on our old wireline network, that turned out not to be very wise. If you allow all your competitors on your network, all services will run on your network, and that results in the lowest cost possible per service which in turn attracts more customers for those services, so your network grows much faster. An open network is not charity from us, in the long run it simply works best for everybody"*.

The access network architecture that best fits the open access models is the Ethernet Point-to-Point (EP2P) architecture, in which a dedicated fiber is deployed to each user enabling a better flexibility in the opening mechanisms at the physical, active and service levels.

In this article we will analyze, through a techno-economic perspective, the viability of the application of open access models to the rollout of

FTTH networks based on EP2P architectures, and compare it with traditional models. To that end, we have developed a specific model in the framework of the COSTA Project (developed by the Polytechnic University of Madrid in collaboration with Telefónica) which calculates the costs related to the rollout of these networks at the different access levels in several geographic scenarios.

Results obtained concerning OANs seem to indicate that a sustainable, profitable, competitive and open deployment can be achieved using this approach.

2 OPEN ACCESS MODELS

The 'open access' term is used to define networks in which different agents are allowed to offer their services over a single network infrastructure, under the same conditions and in a transparent way. The roles that the different operators can play in an open access network (corresponding the different businesses that appear in the value chain) can be represented by a three-layer model. In the passive layer, infrastructure operators connect physically every user to the central office (CO), using fiber. In the active layer, network operators are responsible for network switching and CO's active equipment management. Finally, in the upper layer, service providers tasks can range from traffic aggregation to the provision of innovative services such as telemedicine services.

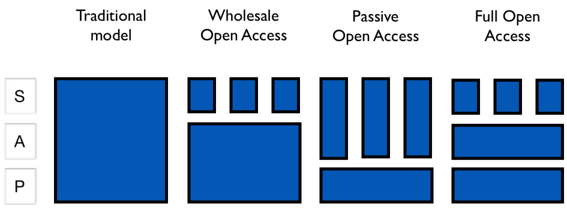


Figure 1: Varieties of open access structures.

Open access networks (OAN) are emerging as accelerators of NGA deployments since they present certain advantages over traditional networks in two dimensions: in the improvement of the viability of the business case and in the reduction of regulatory uncertainty.

Both existing studies (Yankee Group, 2009) and the one that has been carried out in the Costa Project, point out that the key driver for the improvement of the economic viability of a fixed NGA is the ratio of homes connected versus homes passed (HC/HP, take-up rate), due to the strong economies of scale inherent to these networks. Open access networks manage to obtain superior take-up rates as a result of their wider variety of services.

Besides, not only are these networks capable of aggregating a higher amount of services but also experience shows that these services are most valuable for the end user (Ovum, 2009). Service providers have in the OAN an ideal framework in which to develop innovative services, services that truly justify the rollout of NGA's. In Sweden, where a considerable amount of municipally managed OAN already exist, home security, video on demand, remote data storage and online gaming services have appeared and have already become truly differentiating factors. (Informa Telecom & Media UK, 2009).

In terms of competition regulation, open access schemes tries to solve some classic market failures related with vertical integration and significant market power. By ensuring fair, reasonable and non discriminatory terms of access for all players, more competition at service level is achieved easily, ensuring more balanced options for customers. The European Commission has included open access as a mandatory obligation for some types of public aids (EC, 2009a), and in its latest Recommendation, it also encourages co-investment among players, which could be done following these schemes (EC, 2009b).

Despite these major advantages, the implementation of this kind of model may require a tighter regulation in order to fight against the excessive bargaining power that some operators might be able to obtain in their layer, especially in the case of the incumbent. The other big inconvenience falls on that, while it is true that in these networks the global level of competition increases, infrastructure differentiation disappears, slowing down innovation in this area.

The open access model is, nowadays, the approach chosen for the majority of publicly financed networks due to its neutral characteristics. Few operators have chosen it for massive rollouts. However, those who are deploying OAN have relatively high expectations on the project: KPN in the Netherlands, after establishing its joint venture with Reggefiber, is performing commercial tests in 10 cities before scaling, using EP2P following OPTA's recommendation (KPN, 2009). Free, in France, has taken advantage of a favorable legislation to become the second bandwidth operator in the country and has stated to be in favor of deploying an OAN with co-investment (Free, 2009).

Similarly, in Italy, FastWeb, Vodafone and Wind have announced the joint deployment of a FTTH P2P network under an open access model that will reach 4 million homes in 2015, through an investment of €2.5 billion.

3 TECHNO-ECONOMIC MODEL

The techno-economic model used for the performed analysis is based in the Costa Project (http://gtic.ssr.upm.es/costa/costa.html), developed by the UPM and Telefónica (Vergara, 2010). With the intention of studying EP2P open access deployments, we have created a specific model that first considers the deployment of a traditional P2P network and then applies several assumptions in order to study the open access structure.

3.1 Description of the EP2P model

The EP2P model developed simulates the deployment of a FTTH network using a Home Run P2P topology, supporting both Fast Ethernet (FE) and Gigabit Ethernet (GE) connections.

For a standard analysis, a network design to cover all the households of one single area from a central office (CO) for a selected take-up rate is made. The model optimizes network dimensioning for the given service penetration achieving full network occupation of both active elements and cables. The take-up rate can be varied from 0% to 100% in order to generate the cost function for the selected scenario.

Main results are: the monthly cost per user for different take-up rates, total CAPEX, CAPEX per home passed, CAPEX per home connected, CAPEX per customer, OPEX, and IIR or NPV for a period of time under static circumstances.

3.1.1 Architecture

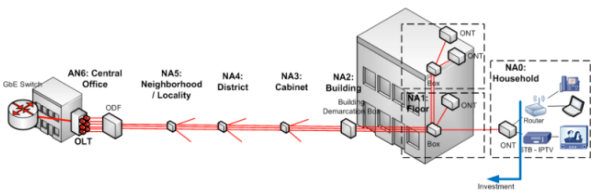


Figure 2: Ethernet Point-to-Point network architecture.

Attenuation and dispersion restrictions have been considered, limiting the maximum length of the CO-user link. Using standard parameters, the maximum lengths obtained have been:

	Attenuation	Dispersion
Fast Ethernet	13.5 km	103 km
Gigabit Ethernet	13 km	25 km

Table 1: Distance limitations for standard parameters.

3.1.2 Input parameters

- Homes density, proportion of unpopulated areas, height of buildings.
- Take-up rate, percentage of FE and GE connections, number of aggregation levels.
- Percentage of assured traffic, concurrence factor.
- Equipment data (power output, sensibilities, attenuations, modularities…)
- Percentage of ducts, trenches or poles; manholes distance; fiber protections.
- Equipment price, life cycles, real estate cost, WACC, O&M costs markup, FE and GE ARPU.

3.1.3 Cost elements

The different cost elements considered in the model are: passive infrastructure (ducts, trenches, poles, ODF...), fiber, active equipment (CO Ethernet switches and transceptors), CPE, aggregation and connectivity costs. Additionally, capital, indirect and O&M costs are included.

3.2 Open Access assumptions

In order to adapt the developed EP2P model to the analysis of the open access structure, each of the cost categories defined for the vertical approach have been allocated to the open access layers according to the following table:

	Passive layer	Active layer	Service layer
CPE	0%	100%	0%
Infrastructure	100%	0%	0%
Nodes	0%	100%	0%
Fiber	100%	0%	0%
Conmutation	0%	100%	0%
Transport	0%	0%	100%

Table 2: Cost categories identification with each layer.

In OAN, operators from different layers must share several utilities (such as the real estate or the lighting, cooling and power supply in the CO) so that the interconnection becomes efficient. In the model, these costs have been allocated following the Spanish Local Loop Access Public Offer (OBA).

The input and output cashflows considered when it came to performing the economic analysis of each layer are shown below:

	Inputs	Outputs
Passive layer	• Dark fiber leasing • Collocation	• Infrastructure and fiber investment and costs
Active layer	• Wholesale bitstream rent	• Active equipment & CPE • Collocation costs • Dark fiber leasing
Service layer	• Retail service provision	• Marketing costs • Sales force costs • Wholesale bitstream rent • Transport costs

Table 3: Inputs, outputs used in the economic analysis.

Since risk and ROI perspectives differ among the three layers, different WACC and period figures have been considered in the NPV and IRR analysis of the passive (10% and 15 years), active and service layers (12.5% and 10 years).

The result of these assumptions is a techno-economic model that enables the individual analysis of the characteristics and economic viability of each layer, in an established competitive framework.

4 MAIN RESULTS

With the purpose of studying the impact of open access models, the results obtained for each of the identified layers will be presented and afterwards compared with those obtained for the vertical model. All the results are referred to greenfield deployments with provision of symmetrical 100/100 Mbps services through Fast Ethernet connections.

4.1 Cost curves for the different layers

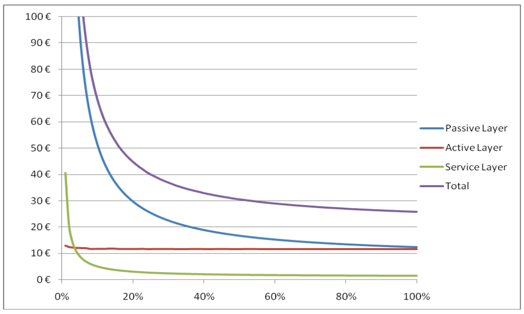


Figure 3: Costs per user versus take-up rate.

The evolution of the cost per user in the different layers relative to the take-up rate offers a noteworthy outlook previous to the study of more concrete scenario, since thee curves will determine the viability of the coexistence of several agents in competition in each level, and whether a stronger regulation should be needed.

4.2 Base case

The base case defined for the presentation of the results considers a 40% take-up rate within an urban area. In relation to the competitive scenario of each layer, it has been considered that only one operator will manage the passive and active layers (usual fact in already deployed networks such as ASTURCON or Citynet Amsterdam) and that there will be three service providers in the service layer.

In the following subsections, we will analyze the economics of the passive, active and service layers, justifying the assumptions taken in the base case and setting the limits of those assumptions.

4.2.1 Passive layer

Economies of scale are the main driver of the passive layer's business case since this layer is characterized by strong investments and variable revenues obtained from the leasing of dark fiber.

The results of the minimum take-up rate needed for NPV to be greater than zero show how rollouts in non-urban scenarios would become non-viable if several agents had to coexist in the layer, since they would not reach enough economies of scale. Even in urban scenarios, the establishment of several operators would increase the uncertainties in a doubtful business case in and of itself. Therefore, the most feasible setting for this layer is the one in which a single operator leases dark fiber exclusively.

Price	Dense Urban	Urban	Sub-urban	Rural	Low Rural
12€	42%	72%	n.v.	n.v.	n.v.
15€	31%	51%	80%	n.v.	n.v.
18€	24%	39%	62%	97%	n.v.
21€	19%	32%	50%	79%	n.v.

Table 4: Minimum take-up rate for NPV greater than zero.

In order to maximize the return of investment at this level, the pricing of the fiber unbundling should represent a tradeoff between the infrastructure operator's short term profit needs and the requirements of the superior layers so as to foster the appearance of innovative services that would, in the end, increase the take-up rate.

It has been considered that a figure of 22€ for the fiber leasing services represents such a tradeoff between both objectives, at least in an urban scenario. However, it has been detected that this price might need to be different depending on the characteristics of the area of deployment.

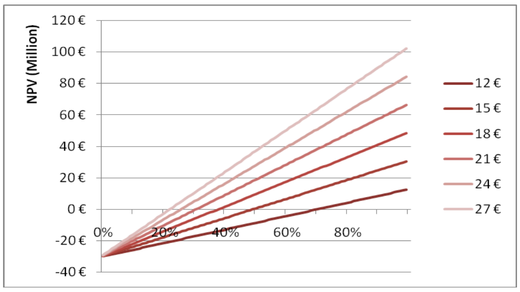


Figure 4: Passive layer NPV for different dark fiber leasing levels.

4.2.2 Active layer

Active operators' business case is less sensitive to the take-up rate since their costs are mainly variable costs with the number of HC. This causes their NPV to be favorable from low take-up levels, but this also implies that they do not manage to convert new clients into profits so directly.

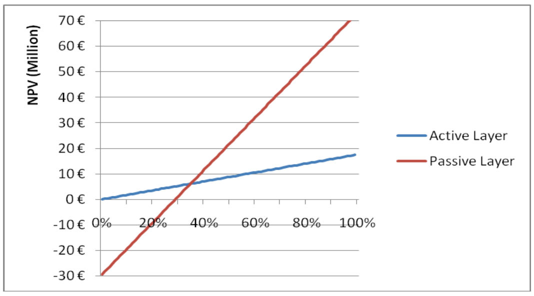


Figure 5: Passive layer NPV vs active layer NPV.

Naturally, an active operator will focus on improving the main drivers of its business case. Since it has been proven that the take-up rate is not one of these drivers for active operators, their actions may not be aligned with the sustainability of the whole network. Thus, regulation in this layer is crucial to avoid abusive wholesale bitstream prices.

Nevertheless, if instead of being offered preconfigured service packages of 100 Mbps, end users were offered single services from which they could configure their subscription, active layer objectives would reconcile with those of the whole network. In such a situation, active operators would have incentives to lower the price of their wholesale bitstream so as to foster innovative services to appear, services that would make end users contract maximum capacity in their lines. Only in this case could free competition work in the active layer.

For this article's analysis, it has been considered that the wholesale bitstream pricing is set to 0.17€/Mbps, although, again, it must be said that this price should vary depending on the characteristics of the area of deployment.

4.2.3 Service layer

Service layer is, without any doubt, the one that will determine the viability of the rollout. The key driver of the general business case is the take-up rate, and it will strongly depend on the variety, innovativeness and quality of services available.

The NPV analysis of the case of several service providers with different market share (over a total take-up rate of 40%), shows how the profits magnitude depends largely in the penetration index, although a low level of the take-up rate does not risk in most cases the economic viability of this layer.

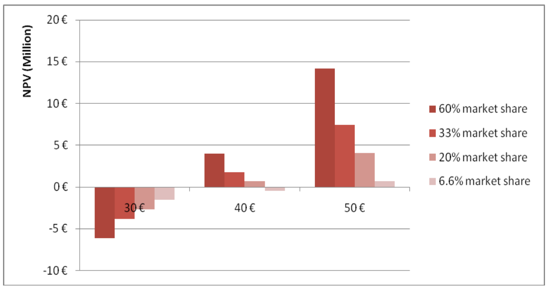


Figure 6: Service layer NPV for different market shares.

This analysis applied to a scenario with three service providers, indicates that the combination of a dominant operator and a stable market could only be obtained for ARPUs over 40€ (having 60%, 20% and 20% share each, or even 60%, 33.3% and 6.6%). For lower levels of ARPU only an equal distribution, of the market (33% share each) might be sustainable.

4.3 Open access vs traditional model

To finalize, both the cases of an operator deploying a network and ruling it by an open access model and one managing it traditionally (with a vertical approach) will be compared.

In the first case, it will be assumed that a single operator will manage the passive and active layers exclusively. In the service layer the owner operator will coexist with two alternative operators, all of them having 33% market share. In order to take into account the higher take-up rates typical of OANs, a 15% extra take-up rate has been considered compared to the traditional one (40% versus 25%).

In the second case it will be assumed that both active and service layers are open. In this case, a 5% additional increase in the take-up rate has been considered (reaching 45%) due to the more competitive framework. In the active layer, the owner operator will have the 25% of the market (assuming that, as it has happened with DSL technologies, the owner manages to keep its market share when opening the network) while it will only capture a 12.5% market share in the service layer.

4.3.1 Solely managed active layer

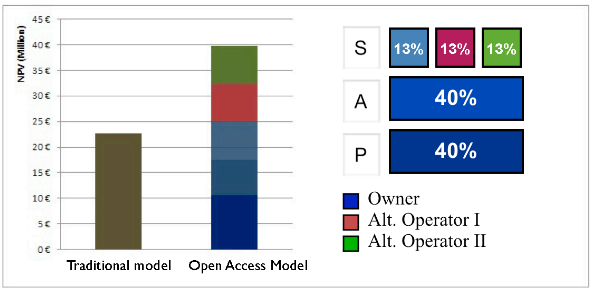


Figure 7: Open access vs traditional approach NPV (I).

Although the assumptions over the increase of service adoption in open access networks have been taken cautiously, results show a better return of investment level for open access networks.

A further analysis, comparing the NPV of the same open access network to the NPV of a scenario with three vertical networks (both having the same total take-up rate), proves the higher stability of the open access market's competitive structure.

Open access scenario		3-player vertical scenario	
Owner	53M €	Operator I	12M €
Alt. Op. I	11.5M €	Operator II	12M €
Alt. Op. II	11.5M €	Operator III	12M €
Total	76M€	Total	36M €

Table 5: Open access vs vertical model: market structure.

4.3.1 Competitive active layer

If, as it was stated in 4.2.2, there happen to be enough incentives in order the free competition in the active layer to exist, the profitability of the business case increases even more.

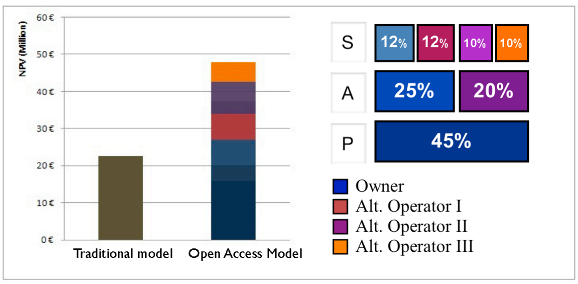


Figure 8: Open access vs traditional approach NPV (II).

5 CONCLUSIONS

The performed analysis indicates that if an adequate pricing is set through the different layers, an OAN can be sustainable and can offer benefits both to operators (because of their improved business case) and to end users (because of the wider range of services at lower price).

The regulator should carry out actions in order to promote the deployment of OANs since they accelerate the rollout of NGA in a transparent and neutral way. These actions include pricing setting in the passive and active layers, which should represent a tradeoff between the operator's short term needs and the sustainability of the whole network and should vary depending on the characteristics of the area of deployment.

However, neither a premature nor an excessive regulation is recommended. All the external actions should be oriented to align the objectives of the different layers with the objectives of the whole network, that is, to increase the take-up rate since it is the most effective way to improving the viability of the general business case.

Results obtained let us glimpse that the main challenge for telecom operators in the near future might be offering a framework in which themselves as well as alternative operators feel comfortable.

REFERENCES

IDATE, 2010. FTTH Global Panorama, status of public funding, *3rd forum national Trés Haut Débit, Caen.*

Noam, E., 2010. Regulation 3.0 for Telecom 3.0. *Telecommunication Policy*, vol. 34, pp. 4-10.

Vergara, A., Moral, C., Pérez, J., 2010. COSTA: A Model to Analyze Next Generation Broadband Access Platform Competition. *Proc. of 14th International Telecommunications Network Strategy and Planning Symposium*, submitted for publication.

Felten, B. Yankee Group, 2009. *Open Access Makes Economic Sense*.

EC, 2009a. *Community Guidelines for the application of State aid rules in relation to rapid deployment of broadband networks*.

EC, 2009b. *2nd Draft Commission Recommendation on regulated access to Next Generation Access Networks (NGA)*.

Informa Telecoms & Media, 2009. *FTTx: A Global Analysis*.

KPN, 2009. Fiber to the home in The Netherlands, *WIK Conference FTTB/H in Europe*

Davis, C. Ovum, 2009. Fibre: the socio-economic benefits. *FTTH Conference 2009.*

Free, Iliad, 2009. *1H 2009 Results & Strategy Presentation.*

Proxecto financiado e liderado pola Deputación Pontevedra

Repoblar

Rede Pontevedra Banda Larga Rural

Vivas onde vivas, sempre conectados

SITUACIÓN INICIAL. METODOLOXÍA. FASES

Total habitantes	Habitantes en Zonas Grises ou Negras
436.281	343.057
% Habitantes con BL	**Habitantes en Zonas Brancas**
78,63%	93.224

EXCEPCIÓNS

VIGO
PONTEVEDRA
CANGAS
VILAGARCÍA
A ESTRADA
REDONDELA
MARÍN

ANÁLISE INICIAL → PROXECTOS PILOTO → SITUACIÓN PARTIDA → CÁLCULO COBERTURA → ESTUDO VIABILIDADE → CONCURSO DESPREGUE REDE → REDE BANDA LONGA PROVINCIA PONTEVEDRA

PROXECTO EUROPEO. ORZAMENTO

Orzamento financiable de 6.693.868,53 €
(76,83 % do solicitado)

Fondos FEDER: 4.685.707,97 € (70%)
Achega Deputación: 1.338.773,71 € (20%)
Achega concellos: 669.386,85 € (10%)

ANÁLISE INICIAL. PROXECTOS PILOTO

- Selección das tecnoloxías máis axeitadas para acadar os obxectivos do proxecto Wimax Rural Pontevedra.
- Consulta ós fabricantes e operadores de telecomunicacións.
- Comparativa Wimax Banda Libre vs. Banda Licenciada.
- Elección de Wimax en Banda Licenciada.

- Catro emprazamentos: Meis, Arbo, Pazos de Borbén e Agolada.
- 7 principais fabricantes mundiais: Alcatel-Lucent, ZTE, Huawei, Cisco, Motorola, Airspam e Alvarion.
- Dous operadores: Iberbanda (3,5 GHz.) e R (2,5 GHz.).
- 20 usuarios friendly por situación.
- Protocolo de probas de servizos e cobertura.

Fabricante	Max Throughtput (canal 5 MHz Beamforming)	Max distancia acadada	Servizos soportados
Alcatel-Lucent	9,3 Mbps DL 2,1 Mbps UL	22 Km. [illegible]	Voz, streaming de vídeo, priorización de tráfico, VPN
Alvarion	12 Mbps DL 2,4 Mbps UL	20 Km. [illegible]	
Cisco	9,8 Mbps DL 2,3 Mbps UL	8 Km. [illegible]	
Huawei	15,1 Mbps DL 2 Mbps UL*	17 Km. [illegible]	
Motorola	11,4 Mbps DL 1,9 Mbps UL	7,7 Km. [illegible]	
ZTE	15,6 Mbps DL 1,99 Mbps UL*	15,5 Km. [illegible]	

DESPREGUE DA REDE. PROXECTO

PROXECTO *REPOBLAR*

Co proxecto **Wimax Rural Pontevedra** preténdese **REPOBLAR** a provincia... rachar coa brecha dixital, de xeito que todos os cidadáns, vivan onde vivan, teñan as mesmas oporturnidades...

Unha provincia máis conectada, máis unida, na que todo estea máis preto, na que todos os cidadáns poidan conectarse...o obxectivo é **REPOBLAR Pontevedra**.

FASES

1ª Fase
5% equipamento
Data límite 31/07/2010

2ª Fase
20% equipamento
Data límite 31/08/2010

3ª Fase
60% equipamento
Data límite 15/10/2010

4ª Fase
100% equipamento
Data límite 31/12/2010

SITUACIÓN FINAL. DATOS GLOBAIS

REPOBLAR abarca 55 concellos da provincia de Pontevedra	Orzamento 5.900.000 euros (IVE incluído)	Adxudicación 5.876.400 euros (IVE incluído)

Poboación total (55 concellos)	436.281
Poboación con acceso a Banda Larga actualmente	343.057
Poboación obxectivo (zonas brancas)	93.224
Poboación obxectivo cuberta	74.219
Poboación non obxectivo cuberta	252.867
POBOACIÓN TOTAL CUBERTA CO PROXECTO	327.086
Poboación con acceso a Banda Larga despois da actuación	417.276

Servizos	Previstos nos pregos	Ofertado por Iberbanda
Servizo de acceso a internet básico 1Mb de baixada e 256 k de subida 20% de caudal garantido	19,95 €	17,00 €
Servizo de acceso a internet avanzado 2Mb de baixada e 512 k de subida 20% de caudal garantido	36,00 €	34,00 €
Cota de alta	100,00 €	59,00 €
Compromiso de permanencia	18 meses	0 meses
Servizo de telefonía	10,00 €	6,00 €
Rede Privada Virtual (VPN) para conectar os concellos coa Deputación 1Mb de baixada e 512 k de subida 20% de caudal garantido	Gratuíta	Gratuíta

UNIÓN EUROPEA
Fondo Europeo de Desenvolvemento Rexional
"Unha maneira de facer Europa"

www.repoblar.depo.es

Deputación Pontevedra

PRIME TECHNOLOGY IN SMART ELECTRICITY GRID
Adaptation of telecommunication systems to electricity grid

Alberto Sendín, Roberto Guerrero, Pablo Angueira, Javier Morgade
Electronics and Telecommunications Department, Bilbao Faculty of Engineering, University of the Basque Country (UPV/EHU), Alda. Urquijo S/N, 48013 Bilbao, SPAIN, alberto_sendin@ehu.es, rguerrero001@ikasle.ehu.es, pablo.angueira@ehu.es, javier.morgade@ehu.es

Keywords: PRIME, Smart Grid, Metering, AMM, AMI, DLMS, Power Line Communications, PLC, narrowband, low voltage, CENELEC.

Abstract: Smart Grids are the future of Energy Systems, allowing the global benefits of energy efficient ecosystems. Expectations on Smart Grids are high, and a powerful telecommunication infrastructure is an essential ingredient to make them a reality. PLC technologies are available to make telecommunication possible in low and medium voltage levels. PRIME offers an open, royalty and patent free, sound communications technology which has to be deployed alongside infrastructure for electricity provision. PRIME has to be adapted to the specificities of a hundred years old electricity network, and a deep understanding of the implications of the different elements composing the electricity grid is needed.

1 INTRODUCTION

Modern life is not possible without energy, especially electricity. Electricity networks inherit a design not far from what Edison knew, and are composed of expensive assets deployed over decades. Thus modern societies are far from an efficient use of the possibilities of the electricity, some of the main reasons being the lack of control over the electricity networks, the absence of services favouring efficient energy consumption, and in general, static structures around energy business.

Nowadays, a change in paradigm is taking place in electricity networks, driving them towards inter-active and customer-centric networks. This will entail many changes in network design and control. These new networks are collectively referred to as Smart Grids, and will make value added services available in energy supply.

Smart Grids are the paradigm of Internet like networks for energy systems. Much of the technology needed to support these changes, is already available but it has yet to be applied on real systems. This is a tough challenge in such a traditional business as energy. This telecommunication infrastructure needs to be easily integrated with existing assets, in a quick and efficient way. Power Line Communications (PLC) technology has the inherent advantage of using existing power line infrastructure and thus minimizes installation costs and speeds up service provision. PLC technology for the purpose of Smart Grids is applied to low and medium voltage networks (tens of kV, and less than 1 kV, respectively). PLC signals are injected in power lines, using them as a communication channel.

There is a very specific narrowband PLC technology called PRIME (PoweRline Intelligent Metering Evolution) which is being proposed as part of the standardization efforts by main European standard bodies. CEN, CENELEC and ETSI have been tasked, by means of Mandate M/441 (issued by the European Commission's Enterprise & Industry Directorate General) to create European standards "that will enable interoperability of utility meters […] in order to allow timely adaptation of […] 'smart metering'").

PRIME is an open, royalty and patent free, sound physical and media access control layers specification, using narrowband PLC in the CENELEC-A band over the low and medium voltage part of the electricity distribution network. PRIME connects energy meters, different gadgets (displays, control boxes, …) and data concentrators with a telecommunication network working over the electricity grid. PRIME offers to these devices, network auto-configuration, self recovery mechanisms, plug&play installation, and all the needed modern mechanisms present in state of the art PHY and MAC layers.

Although PRIME is primarily concerned with AMM in its first implementations, extensions to other functionalities such as Demand Side Management are straight forward. PRIME strongest value proposition is interoperability for different vendors' equipment and systems.

The challenges PRIME is facing are related to the deployment of the PLC telecommunications technology over electricity networks, both in the assets of the utilities, and in the premises of the customers. These aspects have not been studied previously and may prevent the success of PRIME in real massive deployments if not handled properly.

2 THE ELECTRICITY GRID and PLC SYSTEMS

An electricity system is composed of four main components: generation plants (source of the Energy in the system), transmission lines (to transport the energy), substations (to adapt the electricity to the different voltage levels present in the system) and the distribution system.

While transmission lines' voltages are around the hundreds of kV, distribution system lines can be classified as medium voltage (tens of kV) and low voltage (< 1 kV). PLC for the purpose of Smart Grids is applied to low and even medium voltage networks (LV and MV). PLC signals are injected in power lines, using them as a communication channel.

MV networks link together the different secondary substations (where MV to LV transformers are placed). From transformers, LV networks flow as far as the households, with a bus – shared media topology as in the attached figure representing a typical European urban LV distribution network scenario.

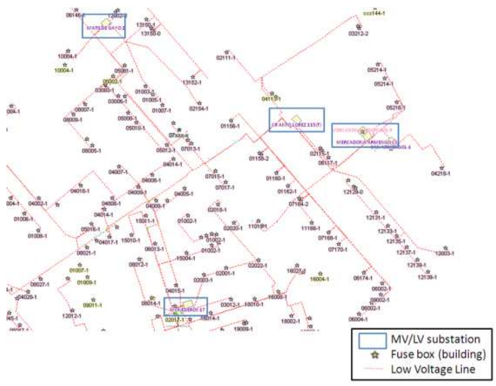


Figure 1: Electricity distribution, low voltage.

However, electricity grids are not the same in different parts of the world. Two different models can be used as a reference:

- European model. MV voltage levels are around 10 kV or 20 kV, with overhead and underground cables: in urban areas, cables are typically three phase, underground, and with transformer coping around 300 households in distances that in urban scenarios reach up to around 200 m. MV to LV transformation usually exhibits a star configuration, with neutral to ground, that provides 400 V between phases and 230 V between phase and neutral. Thus PLC signals can be injected between phase and neutral.
- North American (US) model. MV distribution uses voltages between 4 and 34 kV, with lengths around 15 and 50 km. MV to LV transformation supplies voltages of 120 V or 220 V depending on the load. Typical lengths in LV distribution are up to around 300 m with an average of 10 users per transformer. PLC can be injected between phase and neutral at 120 V.

The intrinsic nature of the electricity grid, made of power lines of different voltage levels, gives the possibility of using them for the purpose of the new Smart Grid. Moreover, the ubiquity of the power distribution grid sometimes renders PLC as the only alternative for data communications.

PLC communications may be used for a variety of services needed (remote control, automation, remote meter reading…). Referring to PRIME in metering applications, this technology is applied in most cases in LV distribution network, from MV to LV transformers, to the different meters that measure the energy consumed in each household.

3 WHAT IS prime?

PoweRline Intelligent Metering Evolution (PRIME) project was launched in 2006 to provide open, royalty and patent free, physical and media access control layers, together with definitions for certain convergence layers, for narrowband PLC solutions operating in the CENELEC-A band (3 to 95 kHz) in the low voltage part of the network.

PRIME technology is intended to provide a telecommunications network over PLC to connect meters to their transformer substation, from which other communication technologies provide links to the control centres.

3.1 System Architecture and Concepts

The basic architecture for PRIME compliant systems is shown in the following diagram.

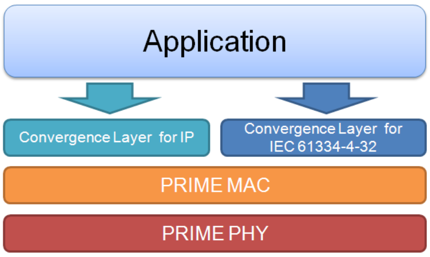


Figure 2: PRIME layers.

PRIME defines the PHYsical (PHY) and Media Access Control (MAC) layers and serves via convergence layers applications such as Device Language Message Specification / Companion Specification for Energy Metering (DLMS / COSEM) but not restricted to metering (IP protocol layer is also supported), which provide a protocol to communicate over the electricity grid. One of the main concepts of PRIME is network auto-configuration. PRIME defines mechanisms at MAC level that allow the network to be dynamically reconfigured at each moment, irrespective of network conditions at any given moment.

3.1.1 PRIME System Concept

PRIME system is composed of subnetworks, which are defined within the context of a transformer substation and its dependent devices, such as meters in metering applications (similar to GSM systems, different Base Nodes with coverage areas initially restricted to its low voltage reachable customers). A subnetwork comprises a "tree" with two types of node, one Base Node and multiple Services Nodes.

The Base Node is at the root of the "tree" and manages the subnetwork resources and connections. There is only one Base Node in a subnetwork, acting as the master node that provides connectivity to it. This Base Node is connected to the LV network in the MV to LV transformer at the secondary substation.
Nodes which are not the Base Node are called Service Nodes (the "leaves" and "branches" of the tree, Terminal Nodes and Switch Nodes respectively). In meter applications, Service Nodes are the PLC modems of meters installed in customer premises. These nodes configure themselves to establish network connectivity up to the Base Node in the PLC network. A Terminal Node is a "leaf" of the tree, while a Switch Node ("branch") is the basis for the connectivity within the subnetwork, by means of the layer 2 "switching" mechanisms implemented. These mechanisms are based on a dynamically created network topology, distributed through the different Switch Nodes, which have a limited knowledge of the network whose functioning they assist (not to require vast amounts of memory in devices that by nature must be low cost).

3.1.2 Network Autoconfiguration Concept

Service Nodes may become either Terminal Nodes or Switching Nodes; conceptually, both apparently different devices are, in fact, states of a same electronic device.

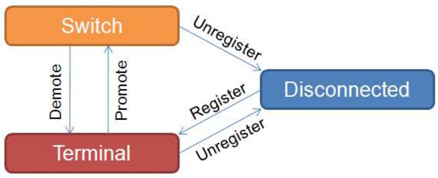


Figure 3: States and transitions in Service Nodes.

The mechanism implemented in PRIME is a consequence of the need to have nodes in the system that could assist the Base Node in reaching more distant parts of the network. A node which cannot directly communicate with the Base Node with just physical layer resources (due to noise or interference constrains) could manage to get its signal to be "relayed" by a Switch Node so configured. Terminal Nodes, speaking of real devices and for metering applications, are basic meters that have to connect to the Base Node; Switch Nodes are plain meters (Service Nodes), acting also as switches.

One of the most interesting features of this Service Node duality is that it allows the network to be dynamically reconfigured whenever there is a situation that requires a certain node to help any other one to progress its data up to the Base Node. It is through control messages that the different nodes configure themselves as needed to set the whole network.

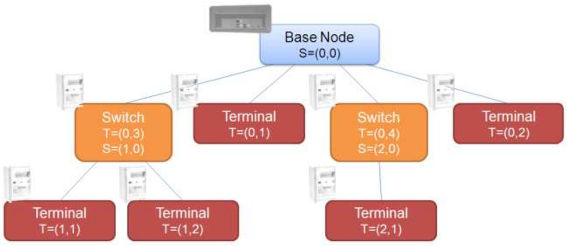


Figure 4: Example of subnetwork "tree".

The attached figure is an example of this situation. The different nodes in the figure include a simplified addressing scheme in which each node has two numbers, and some of them two groups of numbers. The basic idea is that Service Nodes in "switch" state have two pairs of numbers (one address as Terminal Node -T-, and another one as Switch Node -S-). Terminal Nodes exhibit only one pair of numbers, where the first one identifies the relationship to its Switch Node (the one helping him to reach the Base Node), and the other one is the sequential number of the different Terminal Nodes below that Switch Node.

The addressing shown is a simplification of the more general one that operates in the system. This specific addressing scheme identifies subnetworks and different types of node, manages connections, and performs multicast and broadcast.

3.2 PHYsical Layer (PHY) in PRIME

PRIME PHY layer is designed to transmit and receive over power lines which were originally devised for distribution of power at 50-60 Hz AC. The use of this medium for communications at higher frequencies presents some technically challenging problems.

Distribution networks are usually made up of a random variety of conductor types, terminating into loads of different impedances. Such a network has an amplitude and phase response that varies widely with frequency. Furthermore, the channel characteristics can also vary with time as the loads on the network change.

Interference also affects power lines. Electric appliances with different kind of engines, switching power supplies and halogen lamps produce impulse noise that reduces the reliability of communication signals. Due to attenuation, noise is also location dependent.

Thanks to a combination of approaches it is ultimately possible to overcome the hostility of the medium and allow for robust high speed, low cost communications over power lines. PRIME is based on adaptively modulated Orthogonal Frequency Division Multiplexing (OFDM), along with forward error correction and data interleaving. OFDM is very robust against impulsive noise and it achieves high spectral efficiencies, thus allowing for higher data rates.

PRIME works in a frequency subband inside the CENELEC A band ("restricted to electricity distributors and their licensees", as the EN50065-1 states, spanning from 3 kHz to 95 kHz). A total of 97 data subcarriers are transmitted between 42 kHz and 89 kHz adaptively using one of three differential digital modulation schemes (BPSK, QPSK, 8PSK) and optionally a convolutional code.

Table 1 – Raw speed at physical layer.

	BPSK	QPSK	8PSK
Coding on	21.4 kbps	42.9 kbps	64.3 kbps
Coding off	42.9 kbps	85.7 kbps	128.6 kbps

PRIME modems used for the purpose of this paper have direct access to PHY layer resources.

3.3 Medium Access Control (MAC) Layer in PRIME

PRIME MAC layer considers the subnetworks defined by a transformer substation and its associated meters. As previously stated, each subnetwork has a Base Node located in the transformer substation and multiple Services Nodes located in the meters. With regards to MAC, it defines the needed functions and procedures to get the system concept to work.

PRIME MAC layer intends to take the most out of the physical layer below. MAC layer takes inspiration from meshed systems, whilst considering that the available bandwidth will be limited. Simplicity, low cost and flexibility are the main goals of the design.

3.3.1 MAC PDU (Packet Data Units) Formats

Information in MAC layer is structured in Packet Data Units (PDUs). There are three different types of PDU:

- Generic PDU, used for nearly every transmission, except the two PDUs below.
- Beacon PDU, used by Base Node and Switches to announce the subnetwork and its parameters (security profile, identification of the subnetwork and the device generating beacon, Subnetwork Address –SNA- of this subnetwork, and details of frame structure to be followed for the duration of frame that starts with this beacon).
- Promotion Needed PDU, used when no beacon is received, to search for help in reaching the Base Node.

Generic MAC PDUs (GPDU) are the most prevalent ones, and hold most of the subnetwork traffic. They convey data and control information.

3.3.2 Base Node as the Organizer

MAC layer is defined for a connection oriented master/slave configuration. Connections are needed for the communication of nodes in the system.

The Base Node is the node in the system which controls the situation inside its subnetwork, also in charge of organizing the frame structure.

PRIME devices will work in a shared communication medium (low voltage electricity cable), and will use a channel access scheme based on CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance) scheduled along with Time Division Multiplexing.

Time is divided into fixed size frames, and transmission of a beacon (Bc) by a Base Node identifies the start of the frame. There are some beacon slots reserved for Base Node and Switch Nodes transmission (periodicity is flexible), and the rest is shared dynamically between a slot for shared medium access (SCP, Shared Contention Period) and another one for collision free transmissions (CFP, Contention Free Period) for fixed capacity assignment. The frame has a fixed duration, and the SCP has a minimum duration of 64 OFDM symbols time.

3.3.3 MAC Control Packets

Control information of the MAC is sent using MAC control packets. Downlink control packets are sent by every Switch Node (including Base Node) to communicate control information to one of its nodes, or to all of them. Uplink control packets are sent by Service Nodes to communicate with the Switch Node they depend on, or with the Base Node.

There are 10 control messages defined in the system, namely, Registration management (REG), Connection management (CON), Promotion management (PRO), Beacon Slot Indication (BSI), Frame structure change (FRA), Contention Free Period request (CFP), Keep Alive (ALV), Multicast Management (MUL), PHY Robustness Management (PRM), and Security information (SEC).

3.3.4 Retransmissions

The power line channel at PHY operational frequencies can experience different types of noise, among them impulsive noise. While the duration of these noisy impulses is very short, it is bound to cause problems for any data on the channel. For this and other causes, to enable better application layer efficiency, an optional retransmission scheme is specified within the MAC.

The specified ARQ mechanism is that of “Selective Repeat”.

3.3.5 Security

The security functionality provides privacy, authentication and data integrity to the MAC layer through a secure connection method and a key management policy. The network may choose to encrypt or not to encrypt data; Beacon PDU and Promotion Needed PDU are transmitted non-encrypted.

Several security profiles are provided to manage different security needs, which could arise in different network environments. Current version of the specification enumerates two security profiles, namely Security Profile 0 (no encryption) and Security Profile 1 (128-bit AES encryption).

3.3.6 Optimal Network Configuration

The auto configuration possibility PRIME system exhibits, allows many options for network convergence. There are fields defined in Beacon PDU and ALV control packets to get the network to an optimal configuration status.

3.4 Convergence Layer

Convergence Layer (CL) opens MAC and PHY layers to upper layers and applications. CL classifies traffic associating it with its proper MAC connection. This layer performs the mapping of any kind of traffic into MAC SDUs, providing access to the core MAC functionalities. It may also include payload header suppression functions.

PRIME convergence layer is separated into two sublayers:

- Common Part Convergence Sublayer (CPCS) provides a set of generic services, common to any SSCS. For example, the Segmentation And Reassembly (SAR) segments convergence layer SDUs that are larger than a specific size into fixed size segments. The segmented data is reassembled at the destination SAR before being forwarded to applications.
- Service Specific Convergence Sublayer (SSCS) contains services that are specific to one application layer. PRIME defines two convergence layers in order to accommodate different kinds of traffic into MAC SDUs, namely, IP (both version 4 and 6) convergence layer as a very useful and universal access to PRIME, and IEC 61334-4-32 as a simple link towards metering systems.

4 WHiCH are the premises where PRIME enabled devices aRE PLaced?

Focusing our attention on low voltage electricity grid, we find ourselves in the Transformer Centre (TC), from which all electricity signals flow towards customers. Just behind the transformers, we find low voltage bus bars, with the different low voltage lines with different paths towards all the households. There, we find fuses to delimit responsibility in the maintenance of the grid, and just behind them, meters are placed to be able to measure electricity consumption.

Some PRIME devices are installed in TCs, injecting communication signals into the low bus bars. These are the Base Nodes. Their PLC signals flow, alongside with electricity power in the low voltage cables towards fuses and meters, where PRIME devices also exist.

Utilities often exhibit a common structure in the provision of all these elements, although when monitored from a closer perspective, details may be found to differ between different implementations worldwide, and even in different countries and between different utilities in the same country.

Following description will be referred to Spanish case, where Iberdrola will be used as the reference.

The different elements present in the distribution network are:

- Transformer centres (transformers, low voltage bus bars).
- Low voltage lines.
- Fuse boxes.
- Meters

4.1 Transformer Centres

TCs are premises where voltages are transformed from medium to low voltage and / or that allow the easy connection and disconnection of medium voltage lines.

Electricity TCs are also known in electricity world as secondary substations, and even as just transformers.

TCs may be classified following different criteria:

- Depending on how they are connected to the grid (pass through or at the tail of the line).
- According to property (belonging to utility or property of customers).
- According to its placement (indoor or outdoor).
- In accordance with the connection type (overhead or underground).

Elements inside TCs must comply with very strict rules that in the case of Iberdrola are called NI (Iberdrola Norm), such as NI 72.30.00 for three phase oil transformers and NI 72.30.06 for three phase other liquids different form insulating mineral oil transformers.

Low voltage panels distribute electricity, and from them the low voltage electricity cables towards the customers, flow. These panels house low voltage line fuse bases, and they are the interface towards the transformer. Inside the panel, several bars may be found (three phase and neutral), with a group of blade type fuses for each output circuit. All these components are mounted on a metallic framework (old panels) or inside a metallic enclosure (modern panels).

The panels must comply with NI 50.44.02 and the bases with NI 50.48.21. There are three different types of panel: open panels, panels with open bases and metallic enclosure, and panels with closed bases with metallic enclosure.

4.2 Low Voltage Lines

Low voltage lines are linked to the TC, and reach all customers. Low voltage lines are protected by an element to manoeuvre and protect. They transport signals of less than 1,000 V: B1 (0.127/0.220 kV), B2 (0.220/0.380 kV), and B2A (0.231/0.400 kV).

Low voltage lines may be classified according to their hanging or laying strategy: conventional (over poles), plied on facade, plied over poles, clamped on facade, trench, tube and gallery.

Low voltage cable types as defined by Iberdrola, with aluminium being the most typical metal used, are:

- DN-RA 0.6/1 KV AL (1x150, 1x240, 1x25, 1x50, 1x95) underground according to NI 56.31.71.
- RV 0.6/1 KV AL (1x150, 1x240, 1x25, 1x50, 1x95) underground according to NI 56.31.21.
- RZ 0.6/1 KV AL (2x16, 2x25, 3x35/54.6, 3x50, 3x70, 3x95, 4x16, 4x35, 4x50, 4x70) overhead according to NI 56.36.01.
- RZ 0.6/1 KV AL (3x25/29.5, 3x95/54,6, 3x150/80, 3x150/95, 3x150/95+22 LAC) overhead according to NI 56.36.01.

4.3 Fuse Boxes

Points of supply are always attached to house connection boxes. Inside these boxes, there are fuses used to delimit the responsibility of the utility providing the electricity service, and to provide protection mechanisms. These fuses are the last element of the utilities' distribution networks, and the closest to the customer.

House connection boxes must comply with NI 76.50.01. Inside them, the fuses are closed unipolar bases and must be compliant with NI 76.01.02.

4.4 Meters

Spanish regulation has experienced an enormous change with the advent of the Royal Decree 1110/2007, August the 24th. With this mandate, Spain gave the first steps towards the establishment of a remote management system for meters. The characteristics of the meters according to Spanish type 5 (points of supply less than 15 kV) are defined in this legislation, and it is mandatory that they are capable of being remotely managed.

The mandate is not specifying the telecommunication technology this system should be using, although it mentions the different communication media that could be used for the physical support of communications (it includes PLC, but also radio and cable support), and states that the protocols used must be preferably of a public nature.

The order ITC/3860/2007, December the 28th, fixed a very challenging time schedule for the substitution of these meters, and as well as for the deployment of the system level part of it: all meters attached to points of supply less than 15 kW must be replaced by remote management capable meters by the end of year 2018.

5 PRIME SIGNAL INJECTION IN TRANSFORMER CENTRES IN A SAFE ENVIRONMENT

Security aspects related to each and any premise and element associated to electricity supply are of paramount importance. These security aspects, as the ones related to the safety of people and devices, have to be observed by any technology being deployed in electricity premises, disregard of the effect it may impose on the technology.

These safety and security aspects, specifically for the TCs, are necessarily affecting PLC technology being deployed, due to the need to place the devices injecting signal in the electricity grid behind a whole lot of protective elements. The devices providing communications must be capable of bearing the effects produced by these elements, not specifically designed for these PLC communications.

5.1 The Protective Elements

Iberdrola norms prescribe a number of protective elements to be used when connecting devices to low voltage panels in TCs:

- Tetra-polar circuit breaker as of EN 60 947-2; In=16 A; Icu=25 kA; Ics= 75%Icu; D curve (A element).
- Differential tetra-polar circuit breaker 4x25 A; 30 mA; AC type (B element).
- Differential bi-polar circuit breaker 2x25 A; 30 mA; AC type (C element).

5.2 The tests

The tests have been executed with two pairs of devices. These devices are PRIME PHY capable modems such as those inside Base Node and Service Nodes.

One pair has the possibility of injecting PLC signals not through the AC line modems use for their own power supply, and the other pair has the PLC signal injection and the AC power supply on the same copper pair.

The provider of these devices is ZIV company. The devices have been configured to transmit at maximum power, with different frame length (number of bytes), and at all the possible modulation schemes (both coded and uncoded: dbpsk, dqpsk, d8psk, dbpsk_f, dqpsk_f, d8psk_f). These devices are using the physical layer procedures defined by the PRIME Alliance for physical layer compatibility tests.

The table below summarizes the results when the PLC signal is detached from the power supply and when the signal is attached to the power supply, for frame error free communications in all modulation schemes except for d8psk.

Table 2 – Attenuation in PRIME signal injection.

	Atten. (dB), power supply detached	Atten. (dB), power supply attached
C element	0.14	1.39
B element	0.14	1.37
A element	0.49	1.79

As conclusions from these tests we can say that:

- The attenuation introduced by the protective elements is negligible, considering the rest of possible losses in the PLC system.
- The attenuation above is lower when PLC signal is not coupled through the line feeding the modem, so that PLC injecting through cables detached from power supply might provide a better performance in real deployment scenarios.
- Focusing on the errors received in the transmission, and as already anticipated in previous investigations, d8psk uncoded modulation scheme has a low probability of being used, especially in noisy environment with long frame lengths. So this measurement, always with errors in the setup tested, is not considered relevant for the purpose of the study.

6 CONCLUSIONS

Smart Grids are the future path of worldwide Energy Systems, allowing the global benefits of energy efficient ecosystems. Expectations on Smart Grids are high, and a powerful telecommunication infrastructure is an essential ingredient to make them a reality.

PLC technologies are available to make telecommunication possible in low and medium voltage levels. PRIME offers an open, royalty and patent free, sound communications technology which has to be deployed alongside low voltage infrastructure.

PRIME has to be adapted to the specificities of a hundred years old electricity network, and a deep understanding of the implications of the different elements composing the electricity grid is needed. This paper includes the conclusion on the adaptation of PRIME to electricity safety requirements in TCs.

ACKNOWLEDGEMENTS

Authors want to thank all the companies involved in PRIME Alliance, and in particular, authors would like to express gratitude to Iberdrola, ZIV and Current Technologies for their availability for the different tests performed.

REFERENCES

EN 50065-1 Signalling on low-voltage electrical installations in the frequency range 3 kHz to 148.5 kHz - Part 1: General requirements, frequency bands and electromagnetic disturbances.

A. Sendin, 2008, *Iberdrola PRIME Project: Above PHY Layer* Metering International, Issue 4.

I. Berganza, 2008, *Iberdrola Strategy on AMI and Smart Grids*, Metering International, Issue 2.

A. Sendin, A. Llano, P. Angueira, 2009 *Análisis de la utilización efectiva de los esquemas de modulación de la especificación PRIME para la lectura remota de contadores*, XIX Telecom I+D (in Spanish).

Henry Arturo Bastidas Mora, Sep./Dec. 2005, *Modelling some characteristics of electrical networks used as channels for providing telecommunications*, Ingeniería e Investigación, vol.25 no.3 Bogotá.

H. C. Ferreira, L. Lampe, J. Newbury, T. G Swart, June 2010, *Power Line Communications: Theory and Applications for Narrowband and Broadband Communications over Power Lines*, Wiley ISBN: 978-0-470-74030-9.

L. Selander, T. Mortensen, G. Lindell, 1998, *Load profile and communication channel characteristics of the low voltage grid*, , Proc. DistribuTECH DA/DSM Europe 98, London, UK.

Iberdrola, June 2010, *http://www.iberdrola.com*
DLMS, June 2010,
http://www.dlms.com/documentation/index.html
OPEN meter, June 2010, *http://www.openmeter.com*
PRIME Alliance, June 2010,
http://www.prime-alliance.org

SCS (SMART CONSUMABLE SYSTEM)
RFID application in fruit and vegetables packaging industry

Lluis Miquel Marias Albrich
GIRO GH,World lider company in fruit and vegetables packaging industry
Jaume Ribo, 44-58 08911 Badalona (Spain)
lmarias@giro.es

Keywords: RFID consumables. Fruit and vegetables packaging industry. SCS. GIRO.

Abstract: In this paper, it is introducing a new RFID industrial application in fruit fresh and vegetables packaging field, called SCS (Smart Consumable System). In this packaging industry, consumables materials supplied in reels are placed in packaging machinery to produce bags that contains fresh food product. A RFID TAG is incorporated inside the reels core during the process of reels manufacturing and especial RFID equipment is incorporated in packaging machinery. As new features, consumables reels with TAG allows the packaging companies not only controlling of raw material reception and stocks of the consumables used to manufacturing the packages, unless due the information contained in the TAG, is possible to parameterize the baggers according consumables technical parameters, to avoid mistakes between different consumables used to manufacture the package for a specific manufacturing batch, and to have a complete traceability of the consumables used in the bags distributed in the market. Closing the information loop through internet to share the packaging consumables and machinery information between consumable manufacturing company and packaging company, this RFID application allows a considerable progress in packaging process of fresh food suppliers companies.

1. INTRODUCTION

It is well known the Radio Frequency application (RFID) in stocks control of stored products, goods dispatch and reception from manufacturing companies to distribution hubs and internal manufacturing applications to capture process information (1) (2).
There are manufacturing inner applications where TAGS are reusable and therefore their cost it is not the most significant variable to consider. In various industrial process Active tags are used to store the different incidents occurred during manufacturing operations, controlling and monitoring the product quality.

Now a day passive TAG cost is coming more competitive and it is possible to be a part of not expensive products.
Current applications used to control product inside a pallet and passing it trough a set of gantry-type antenna to detect the TAGS and also other applications of product units as bottles or clothes pieces which incorporates passive TAGS, which are detected by portables readers, are well introduced in the market.
There are in the market a divers offer of RFID readers (fixed, mobile and gantry-type) that are available to be installed in different layouts and work shops and ready to work in different application, connecting these dispositive to customized computer system by a specific middleware and if required connected to company informatics system in order to access to corporative BD, but these solutions can not be properly integrated with specific machinery and required row material according the standards of manufacturing processes in packaging industry.

We are going to describe a RFID technology application, named SCS, on industrial intermediate products used in packaging manufacturing by different companies involved in fresh fruit product supplying chain. We are going to explain the integration of this technology in their manufacturing machinery and how is achieving a cost-effectively responding to current problems in their manufacturing processes.

2. PACKAGING MATERIALS SUPPLY CHAIN

Main characteristics of packaging types used in fresh fruit industry are:

- Packing product must have a complete transpiration as it was in free air
- Weights margins of packing product are from 0,5 to 5 Kg
- Product must be seen, smelled and touched by the consumer
- Quantity of material used as packaging must be as small as possible.

Figure 1: Different types of fruit packaging

These features makes that packages for this industry are formed by Knitted or extruded mesh and printed bands (film strips) , besides auxiliary elements like labels and handles.

These types of packaging have been elaborated through different plastic row materials (polyethylene, polypropylene, polyester, bio plastic, etc) depending pack characteristics required as compostable, brightness, welding temperatures, resistance, aspect, thickness, surface communications and enhancement of food color.

The manufacturing process of these consumables has three industrial steps that correspond with three completely different manufacturing technologies plastic extrusion, textile and printing.

Plastic row materials are obtained from mixing different plastic components, supplied in pellet, through cast or blow extruders to obtain bobbins from different materials, thickness and width according to obtain features of the required packaging.

From these plastic bobbins and using flexographic printer machinery, they are obtained reels of printed bands, where printing ink is located between two glued bands (frontal band transparent and back band of solid color, usually white). This printing method assures that ink it is not in contact with food.

Also from other types of plastic bobbins and using special technical textile technology it is obtained tubular knitted mesh reels. Materials, colors, thread size, thread quantity, tubular width, mesh design geometry are some of the parameters that allow to obtain the required mesh.

Auxiliary elements that are parts of a package like handle and thermal labels bobbins also have their specific manufacturing process, based in a extrusion process in the case of the handle and die cutting paper with thermal resin layer in case the labels.

Up to 5 types of these reels it is elaborated a bag:

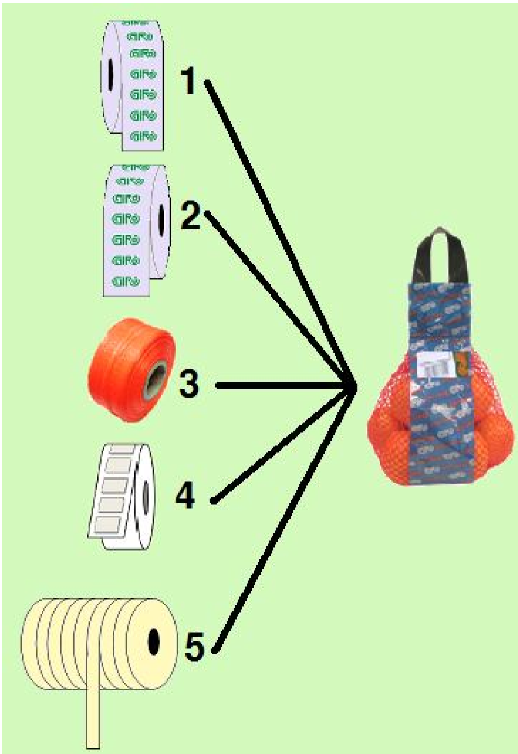


Figure 2: Different types of consumables reels involved in a fruit package

Figure 2 shows the different reels of consumables used to form a bag:

1. Band reel frontal face
2. Band reel back face
3. Mesh reel
4. Label reel
5. Handle reel

Figure 3 shows the automatic bagger machine that forms the bag through the different consumable reels showed in figure 2.

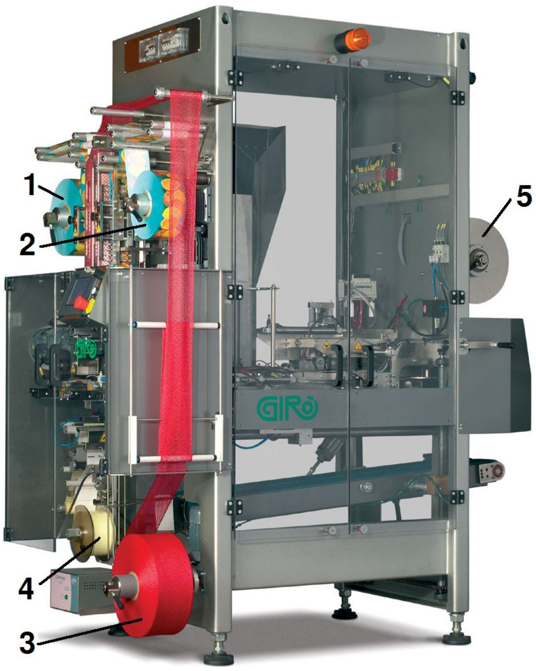


Figure 3: Packaging machine by GIRO (3).

The mesh packaging supply chain starts with chemical companies that manufacture pellets of plastic materials and inks to print plastic bands, second step is made by a company that transform this basic materials in reels of intermediate product and manufacture packaging machines that uses this consumable reels as we are describe in this paragraph, third step is the company that using referred consumables and packaging machinery, packages fruit and vegetables supplying to the retailers and supermarkets, and finally arriving this packaged product to the consumer.

In this paragraph we are summarized the second step process made by the company manufacturer of machinery and consumables as a part of entire supply chain involved in packaging of fresh produce.

3. PACKAGING PROCESS

3.1. Current working process

In this paragraph we are going to explain some existing problems in current working process made by packaging company.

3.1.1. Machinery settings

Packages are formed by different type of bands and mesh that need different machine settings in order to perform correctly the package (welding temperatures, welding times, band length, register type, bag format, label type, handle type, etc…). If these parameters are not set correctly packaging process presents anomalies of different levels that produce incidents in packaging manufacturing process, increasing the cost of not quality. As an example is the operator program in the packaging machine a temperature for a polyester band and in fact the band reel mounted in the machine is polyethylene, the package will present holes in the band showing a bad quality and even resistance lack and package breaks producing the removal of detectives packages and therefore material waste and a delay in delivery time of the order. In fresh product industry due the goods are perishable, supermarkets makes orders according a strictest reposition date of his sales lineal, and delivery time delay makes sales lost in the supermarket.

3.1.2. Consumables selection. Matching

The number of articles used by a packaging company can arrive to thousands, due mainly because bands and labels are customized for every product and customer.
A certain package for a retailer can have slightly different band right and left, also label can be difference for other package of the same retailer in only a little detail difficult to assess, each label has a different printing for each package, mesh ca be supplied in different colors and parameters, etc
Now a day when happens a mistake in the selection of one consumable (mismatching) or selecting a wrong label printing, is very difficult that a operator can detect some little difference between right and wrong consumables, thousands of defective

packages can arrive to the customer causing not only a big cost to the packaging company but a lost of confidence in his quality process. This problem may cause a supermarket not have available this article during several days.

3.1.3. Traceability

As we have already mentioned supplied chain materials has following process:

- Company1. Chemical company who manufactures pellets
- Company 2. Manufacturer of consumables and packaging machinery.
- Company 3. Fruit packaging and supplier.
- Company 4. Retailers and Supermarkets
- Final consumer

Fresh food industry requires to the companies involved in the supplied chain a traceability system, which allows identification of his product batches and their relation with row material batches and packaging materials in direct contact with the food. Traceability system will include all relevant registers of manufacturing and expedition of the packaging product. (4)
Fruit packaging company has a rigorous traceability system of the packaged fruit but it is really difficult to find some one with a correct control of the packaging consumables due the cost required to pick up data from consumable manufacturer and matching with his manufacturing batches.

3.1.4. Shipment , reception, stock and picking process

Actually, once the consumables reels are manufactured, by the consumable manufacturing company, and packaged in 2, 3 or 4 reels by box, pallet with boxes are delivered to the fruit packaging customer. Once arrive to packer warehouse the boxes with consumables reels are stocked in shelves until a manufacturing order requires a picking process to pick up the right consumables. All these working process, including control stock consumables in packer warehouse in order to send consumables reposition orders, are manually.

3.2. Application RFID in packaging systems

3.2.1. RFID system selection

According the target to solve the problematic described in preceding paragraph and complying standards we had to meet the following requirements:

- TAG chosen must comply with EPC global Gen 2 system
- TAG must have user memory enough to store technical data required in machinery settings
- TAG must be assembled in reels core and matches with frequency band selected.
- Antennas must be assembled in packaging machinery assuring not difficult the use of the machine and be designed against damaged in working operations.
- Antennas must be designed to read its associated TAG, but not the TAGS incorporated in neighbour’s reels, considering the major mechanical equipment located around.
- Antenna’s reader must be designed to read until 4 antennas that can be installed in packaging machine.
- Reader must communicate with packaging machine PLC in real time during working process according an adequate protocol.
- Information from TAGS must be included in manufacturing batches data using a middleware that allows storing this information in a BD of a manufacturing plant PC.
- In order to give a integral service to the customer the information from TAGS can arrive to machinery technical services thought internet.

3.2.2. Tag selection

According referred performances we have selected the ALN 9654 G inlay from ALIEN which main features are:

- Meets EPC global Gen 2 (v 1.2.0) as well as ISO/IEC 18000-6C
- Worldwide operation in the RFID UHF Bands(860-960Mhz)
- Higgs™-3 IC with 800-bits of Non Volatile Memory
- 96 EPC bits, extensible to 480-bits
- 512 User bits
- 64-bit Unique TID
- 32-bit Access and 32-bit Kill passwords

ALN 9654G is based on Alien's advanced Higgs™-3 UHF RFID IC and a high-performance antenna allowing to achieve a good communication with the antenna reader designed in the environment where go to work.

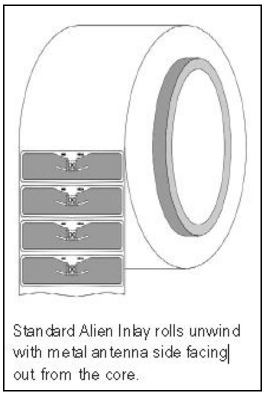


Figure 4: ALN-9654-WR (Wet Inlay)

This TAG has a flat frequency response in world band operation and a radiation diagram showed in Figure 5 that it has been in account in the positioning of the TAG in reel core versus the antenna design and position.

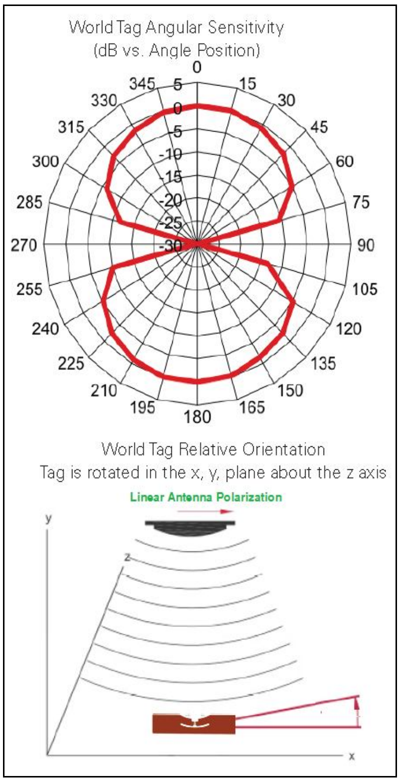


Figure 5: ALN-9654 Inlay Sensitivity & Orientation

3.2.3. Consumables reels manufacturing

The different industrial process to manufacturer bands, mesh, labels and handles reels have in common the reels support core. These cores are made by cartoon or plastic of different thickness depending of the pressure wound.

As we can see in paragraph before the best position to fix the TAG in order to avoid position zones of the reel with shadow is showed in Figure 6.

Figure 6: TAG location in reel core.

TAG record process realized during consumable reels manufacturing are described in Figure 7.

Step A: Operator reads, trough code bar reader, the manufacturing batch from manufacturing order list, and it generates the following informatics process:

Step B: This batch number is used by Middleware application installed in manufacturing plant PC to require product code and technical data, of the consumable reels, to de ERP trough a stored procedure call, using OBDC format.

Step C: Middleware application generates a SQL BD in the plant PC where are stored serial numbers used by every manufacturing batch.

Description	N. bits	Values	
Header	8	0011 0000	EPC (SGTIN-96)
Filter Value	3	000	
Partition	3	101	
Company code	24	7 Dig. GIRO code = 8433084	
Log. Variable + Article code	20	6 dig.: 0 + Product code	
Serial Number	38	Batch number + serial number	

Table 1: EPC structure data

Step D: All the information is structured according the specific protocol used by Zebra printer in order to record RDIF data (EPC (see table 1) + User Memory) and printing on inlay paper data for consumable reel visual identification

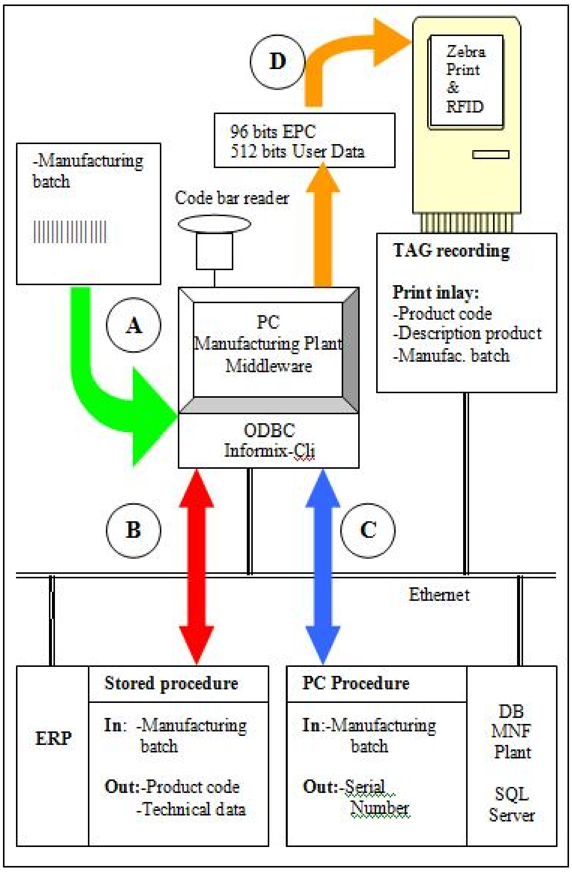


Figure 7: TAG recording process

Once manufactured, consumable reels are introduced in boxes and these boxes are placed in pallets for shipment. With RFID, pallets are passed trough a gantry antenna assuring that shipment product corresponds with manufacturing order and ERP knows in real time when an order is ready in product shipping area.

Depending of consumable type, manufacturing orders are performed to store, these are the case for mesh and handles, in these cases controlling inputs and exits from the warehouse it makes possible to have an updated inventory of the stock. It is easy consider the manufacturer date in pick up orders and assure that oldest product leaves first the warehouse, namely working in FIFO system.

3.2.4. RFID Infrastructure in Packaging Machinery

Once the consumables with RFID arrive to packer warehouse, it is possible to use RFID to make a warehouse process controlled by the information contained in the RFID. This means, stock control in real time, orders management to the consumables suppliers and picking process without mistakes.

Referred process using RFID are well developed and explained in other papers, therefore we will centre our exposition about how to introduce RFID technology in packaging machinery in order to solve the problems exposed in paragraph 3.1. Figure 8 illustrates RFID infrastructure in packaging machinery.

First point to design are the antennas (1 to 4 in the block diagram), located around reels supports in the packaging machine. An antenna switcher controlled by PLC allows entering required antenna signal to the reader. Reader communicates trough RS 232 with the PLC. Packaging machine PLC communicates with PC situated in manufacturing plant and with an appropriate middleware manages all the information with plant machinery. Finally manufacturing plant PC, called GirControl, is connected to ERP trough customer network using Ethernet, and by other side trough Internet, GIRO Technical service is on line connected with customer packaging machinery in order to solve in real time machinery incidents.

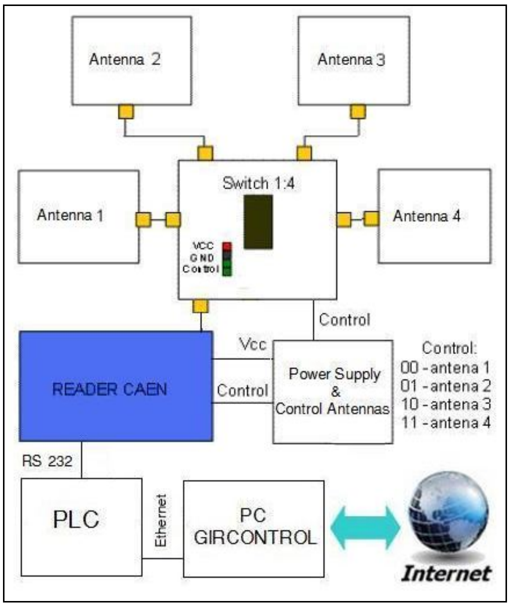


Figure 8: Block Diagram RFID infrastructure

Once selected the TAG and therefore the Frequency band the antennas design must achieve the condition to read the TAG corresponding to selected antenna without interference with the others TAGS situated in the different reels support around the packaging machine, also consumable reels situated around machine (more that 1 meter) not must be detected by a specific antenna. Due that we are working in a industrial plant with a huge electric material, the system must be designed with high noise immunity. Due we are working in UHF frequency, one antenna tested was a type patch, although radiation diagram is adequate, this is an antenna with narrow band width and as consequence of metallic environment the TAG response was shifted and signal was not detected. This antenna design has been rejected for this application.

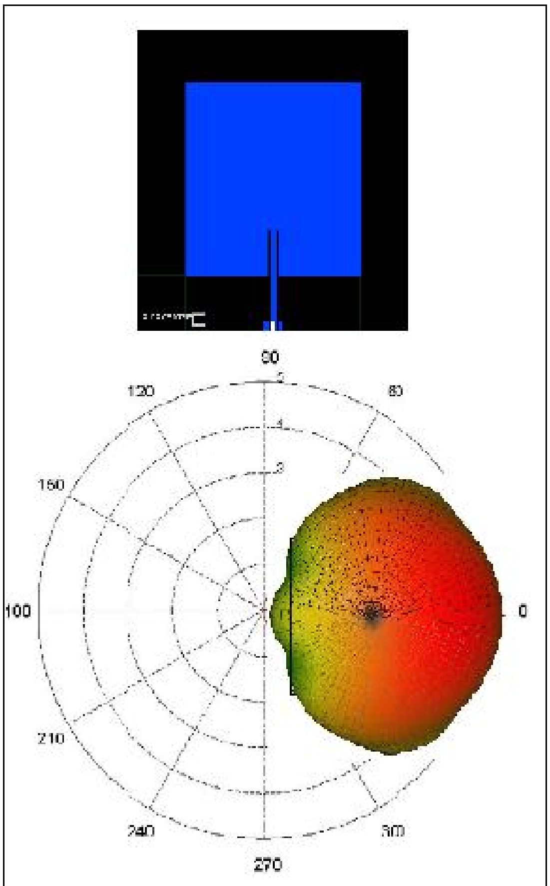


Figure 9: Patch antenna and horizontal radiation diagram.

The antenna selected has been a dipole type with lower gain than a patch antenna but a band width above 100 Mhz. (VSWR<=2 (S11= -10dB's).
Figure 10 shows antenna shape and gain diagram.
Due the industrial environment it has been designed a slot dipole instead a current dipole, with micro strip supplying to assure a balanced dipole with a supply not balanced in the same antenna plane.

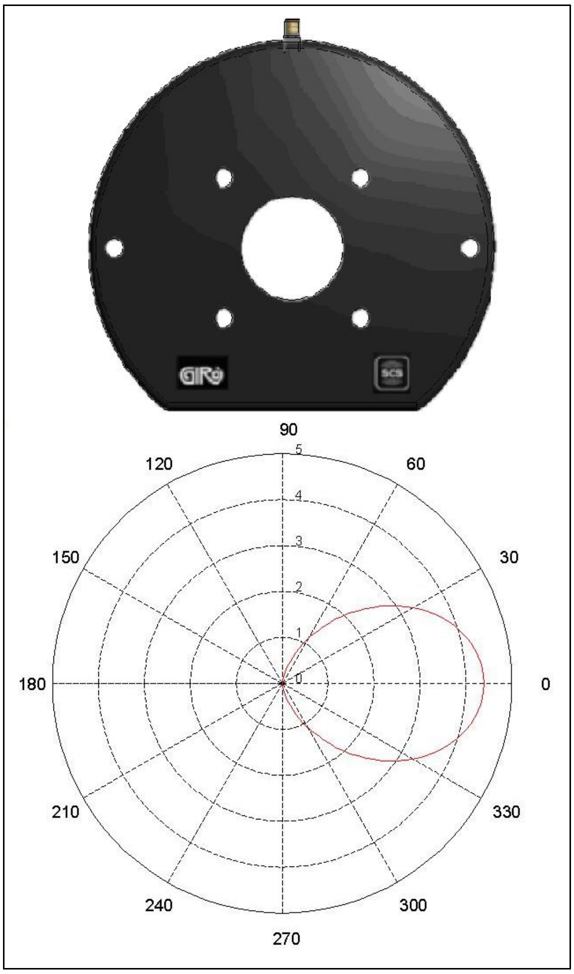


Figure 10: Slot dipole antenna and horizontal radiation diagram.

Depending power level supplied to the antenna it exits 2 reel position zones where antenna can not detect the TAG.
These zones are positioned in upper and lower position of the TAG respect the reel core and the size of these shadowed zones are up 20 º depending materials of consumables reels and a power supplied to the antenna about 200 mw. This Power is adequate to not detect consumable reels located around the machine (more then 1 meter).
Figure 11 shows the antennas position respected supports of consumables reels in packaging machinery.

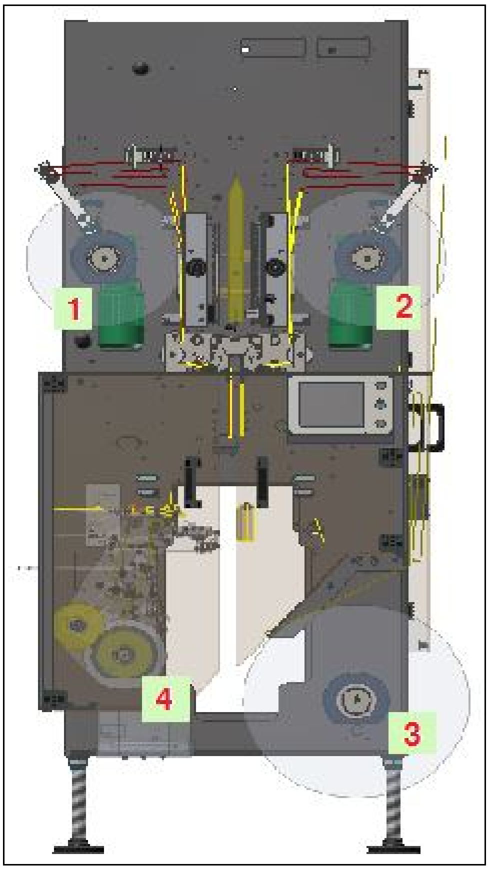


Figure 11: Antennas position coaxial with reels support shaft

The antennas reader selected is A528 OEM UHF multiregional compact reader based on Intel® RFID Technology from CAEN-RFID (5).

Technical Specifications Table	
Frequency	902÷928 MHz (FCC part 15) 865.600÷867.600 (ETSI EN 302 208)
Output Power	Programmable in 8 steps up to 500 mW @ 5 V (27 dBm)
Antenna Connector	Nr. 1 MMCX type
Frequency Tolerance	±10 ppm over the entire temperature range
Number of Channels	10 channels (compliant to ETSI EN 302 208) 50 hopping channels (compliant to FCC part 15)
Standard Compliance	EPC C1G2
Digital I/O	Four I/O lines 3.3V out, 5V tolerant
UART Serial Port	Baud rate: 115200 Data bits: 8 Stop bits: 1 Parity: none Flow control: none 3.3 V out, 5 V tolerant 9.6÷115 kbit/s data rate (settable)
USB Device Port	One USB 2.0 Full Speed (12 Mbits per second) device port.

Table 2: Specifications A528 from CAEN.

This reader has been modified by CAEN in order to adjust the communication protocol to the needing of the application (PLC protocol, scanning timing, and switching power) and it has been added a switching system to select the required antenna (6).
This especial application of RFID in mesh packaging machinery is patented (7), (8).
The position of the reader inside the machine, the type (low attenuation) of coaxial cable used and cables routing inside the machine has been an important task to avoid interferences due the electrical noise caused for existing power lines.
Bagger PLC and PC GirControl operation are going to be explained in following paragraphs.

3.2.5. RFID Process in packaging machinery

Working procedure with RFID consumable reels is the following:

- Bagger is stopped waiting operator charges consumables corresponding to a new manufacturing batch. PLC is switching every 1 second antennas asking to the reader for EPC field of the TAG.
- Operator inserts trough bagger touch screen a new package number, (see figure 12).
- Bagger PLC connects with PC GirControl, asking the consumables codes associated to the package number introduced, writing these codes in the upper position of double boxes.
- Operator insert consumable reels in every shaft support (from 1 to 4, see figure 11). Bagger screen shows every new detection in lower position of double boxes and indicates in video inverse if code no matches with specified by BD GirControl.
- Once charged all consumables, operator starts the bagger. Bagger will go on working if adequate consumables codes are detected during the first three bags (remember that there is a shadow angle that has the possibility that when starting reel is situated in this position, and although the reel always is detected during consumables charging , in this way it is not possible to cheat the system)

In this case if matching is achieved no possible mistakes can happen with the consumables selected to ordered batch.

As we have explained in before paragraphs, there are different packaging types that use different materials, other direct application of RFID system is to check if bagger settings are adequate to the consumable used.

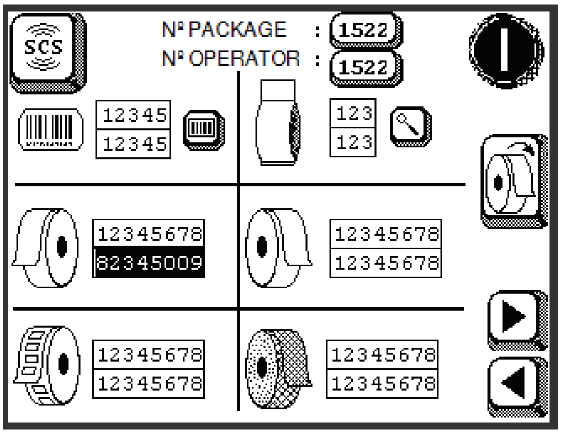


Figure 12: GIRO Bagger screen. Matching consumables

- Once bagger detects a new consumable code, PLC asks to the reader the user memory contained in that TAG.
- If the settings programmed in the bagger not match with the information contained in the TAG, PLC shows an alarm message on touch screen, and shows in a specific screen the different settings between selected and allowed, (see Figure 13).

In this case operator can not make a mistake selecting a wrong machine setting

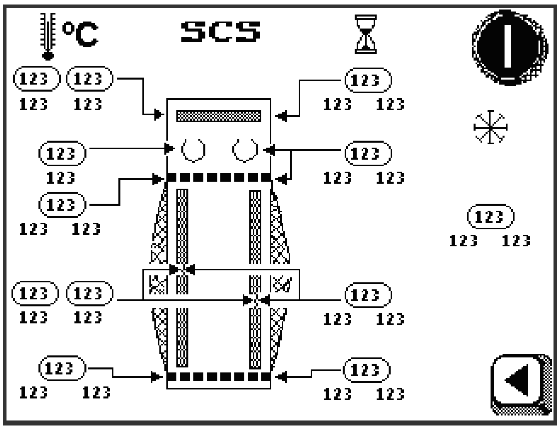


Figure 13: GIRO Bagger screen. Technical data consumables

Finally, manufacturing batch traceability is established because PC GirControl stores for each batch the consumable traceability codes send by bagger PLC.
Hereafter we expound how PC GirControl with a Middleware software, named GirControl, manages the information in manufacturing plant, connecting packaging plant machinery to consumable and machinery supplier company.

3.2.6. Middleware. PC GirControl

There are dozens of baggers installed in a packaging plant and if we add other machinery involved in the packaging process, like weighers, fruit regulators, sizers, etc, it can be installed more than one hundred packaging machinery.
Although it is not intention of this paper to deep in the plant machinery network and the use of GirControl system as manufacturing control system inside manufacturing plant and also, probably the most important feature, to connect in real time trough internet the packaging plant machinery with GIRO technical service in order to get maintenance information and optimize the machinery working, we will introduce GirControl system in order to explain storage traceability information of manufacturing batches.
Figure 14 shows baggers connexion with PC GirControl and connexion to internet.

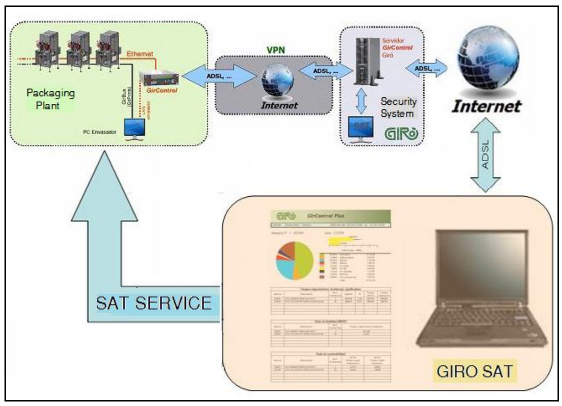


Figure 14: GirControl infrastructure.

In packaging plant, Baggers are connected through Ethernet to PC GirControl, where a routine in C++ following a specific protocol interrogates the baggers in scanning mode, and generating SQL sentences introduces manufacturing batch data in a PostgreSQL BD from PC GirControl.
New packaging machinery capable to access in native SQL writes PostgreSQL BD directly using a JDBC driver.
In plant PCs network user information is showed in WEB page format.
GirControl is an embedded system based in a TOMCAT server that using JQuery framework and SVG graphics serves real time GirControl WEB pages.

PC GirControl can only go out from packaging plant trough a VPN to GIRO server, where a security procedure assures that only authorized people can access.

GirControl can creates by itself the matching table , Nº Package with codes consumable associated (see Figure 15), or trough WEB services routines to inquiry the information to customer ERP.

Figure 15: Matching consumables table

GirControl BD contains all the information corresponding to a manufacturing batch including consumable data that assures a complete traceability, (see figure 16).

Figure 16: Traceability information in a manufacturing batch

4. CONCLUSIONS

SCS is a RFID application for fruit and vegetables packaging industry, which solve different problems that happens in current working process explained in paragraph 3.1.
Try to solve these problems with usual procedures is too expensive and in practice no-viable.

Although this application is designed for a specific industrial sector, it is not known an industrial manufacturing application in other industrial sectors with this scope inter-companies.
SCS matched with GirControl must produce a technological advance and business process improvement in fruit and vegetables packaging sector.

ACKNOWLEDGEMENTS

I would like to thank the good work realized by the team involved in this project:

- AIDACENTER. http://www.aidacentre.com
- DICOEL Electronics http://www.dicoel.es

And especially to GIRO News Product Team, from Company Managers to excellent engineers that make possible to achieve the pursued targets.

This project is part of the effort in R & D carried out by GIRO COMPANY to further advance business development worldwide in the fruit and vegetables packaging sector.

REFERENCES

(1) Sensors d’identitat RFID: Un DNI per a cada producte, TELECOS nº30 3/2005
(2) Proyectos RFID, AECOC- RFID Magazine
(3) http://www.giro.es
(4) IFS International Food Standard (Standard for conducting audits of food products)
(5) RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification, Second Edition Klaus Finkenzeller 2003 John Wiley & Sons, Ltd.
(6) CAEN Technical information manual. Version August 2008 EASY2READ_PROTOCOL_REV8.DOC
(7) Patent EU No. 10382003.1, application date 12-01-2010 title: "Machine for the manufacture of bags"
(8) Patent EE.UU. No. 12/693138, application date 25-01-2010 - title: "Machine for the manufacture of bags"

REBECA: A 3D PROGRAMMING LANGUAGE ENVIRONMENT FOR RAISING ICT VOCATIONS AMONG YOUNG STUDENTS

David Miraut, Ángela Mendoza, Susana Mata and Luis Pastor
Computer Technology and Architecture, Computer Science and Artificial Intelligence Department, E.T.S.I. Informática, Rey Juan Carlos University, C\ Tulipan S/N, Móstoles, Spain
david.miraut@urjc.es, angela.mendoza@urjc.es, susana.mata@urjc.es, luis.pastor@urjc.es

Keywords: Visualization, 3D Animation, Introduction to Computer Science

Abstract: "*Rebeca through the looking glass*" is an educational innovation project based on *Alice* –an initiative from Carnegie Mellon University– that tries to approach and to facilitate the teaching of object-oriented programming to young students. Its multilingual interface is an intuitive and visual 3D environment, designed to motivate their curiosity and encourage them to study ICT (Information and Communication Technologies) careers, without having to cope with frustrating syntax errors or enigmatic core dumps. This paper describes the keys of this new development, how it can be adapted to any language or region, and our experience in the deployment of *Rebeca* in Educational Centers.

1 INTRODUCTION

Enrolment in Computer Science and Telecommunication Engineering majors has been steadily declining during the last years in Spain (and many other countries), in spite of the fact that ITC is a promising sector, full of work opportunities.
Nowadays, high school students are in close contact with ICT. During their education, they even develop some ICT skills, usually through an informal training in these areas. But -paradoxically- they are more reluctant than ever to take an education in ITC studies. In some cases, this behavior can be due to a misperception of the creative possibilities of ICT jobs. For other students, that might happen because of the reputation of ICT-related degrees of being very demanding. Both of these ideas are reinforced by the steep slope of the learning curve during the first year of College: most students experience difficulties with the development of abstract thinking, and they get discouraged when they don't see immediately the applicability of the knowledge they learn in object oriented programming courses.
The Bologna process brings opportunities that must encourage us to look for new teaching strategies which motivate our students to learn on their own from their very first stages at University.
To overcome some of these problems in programming courses, we have developed a new tool, *Rebeca*, based on Alice, an open source software developed by Carnegie Mellon University. *Alice* (Dann et al., 2006) is an innovative tool designed to teach introductory programming concepts to undergraduates with no previous 3D graphics or programming experience. Both *Alice* and *Rebeca* communicate effectively with young people in the visual language

they are familiar with, using a 3D graphics programming environment with a drag and drop interface, where they can make 3D animations and interactive video games while learning to program easily. This is very attractive to students, because the interface is designed to avoid having to face frustrating details related to the syntax of object-oriented programs in their first steps of the learning process. Each object in the virtual world is an object whose behavior can be programmed in a language similar to Java.

Alice has been very successful in the English-speaking countries, but it has hardly been used abroad, because of its lack of support to other languages besides English.

The language barrier is the main stumbling block to introduce *Alice* to young audiences who do not master the English language. Therefore, we focused our efforts on solving this problem and on extending the original software in order to localize it and to make it easily adaptable to any other language. These modifications have required a radical change in the source code and in the programming language itself. This paper describes the keys of the new development, how it can be adapted to any language or region, and our experience in the deployment of *Rebeca* in Educational Centers.

Teaching guides, including a collection of problems and solutions, have been written in Spanish with Creative Common licenses which may also be translated to other languages to exploit the potential of this tool, and to show aspects of our profession to young people in an attractive way.

Figure 1: *Rebeca* splash screen

The following section discusses some of the main reasons behind the decline in ICT studies enrollment. Section 3 describes some initiatives from different universities to change this trend. The following section details the development of *Rebeca*. Section 5 draws its initial deployment in High Schools. And finally, some conclusions and future work to improve this tool are presented in the last section.

2 decline of ICT STUDIES enrolment in Western countries

National Institutes of Statistics, like the U.S. Bureau of Labor Statistics (Bureau, 2010), are clearly optimistic in relation to ICT business opportunities. According to their studies, ITC engineering is one of the sectors that will experiment higher growths during the next decade (44% versus the 11% average growth projected for all industries combined), with excellent remuneration and prestige in specialized positions. However, this increasing demand for ICT engineers is not tuned to the present trend of career choices among young people. This is a real concern both in

Academia and Industry in Western countries.

Since the dot-com bubble burst in 2000-02, fewer and fewer students have been enrolling in technological degrees such as Computer Science, Computer Engineering, Information Technology, Information Systems, Management Information Systems, and Software Engineering (Lomerson & Pollacia, 2006) (Denning & McGettrick, 2005) .

The UNESCO's Institute for Statistics shows in its reports that many European countries (like Spain, France, Germany or Italy) suffer from a deep decline in engineering enrolment, especially in ICT careers, being the number of new students less than 50% with respect to 10 years ago. These figures are similar to U.S.' ones; the National Science Foundation, Computing Research Association and several independent surveys show that enrolment has dropped more than 60% in the same time period (UNESCO, 2010) (Vesgo, 2008).

Employers are concerned about the lack of graduates to fill open positions. This leads either to hire under-qualified personnel or to outsource part of the development to other countries. The first solution implies a decrease in product quality and efficiency, having a direct impact in the company results. Regarding the second one, ICT and Computer Science involve much more than simply "coding". Firms are aware of the fact that the critical functions for the development of their product or the business enterprise's core competencies, as well as those that are critical in maintaining a competitive advantage in the marketplace, are difficult to outsource with success. Although certain IT jobs can be outsourced, the strategic top IT jobs will remain at the company headquarters, simply because they require detailed knowledge and analysis of the key factors specific to a given company or industry (Denning & McGettrick, 2005) (Frieze, 2005)

Many assumptions and conjectures have been made about the reasons that led to the decline in CS enrollment in the early 1990s and again since 2000. Most of the published studies (Lomerson & Pollacia, 2006) (Lenox & Woratschek, 2008) (O'Lander, 1996) agree that one of the main reasons is the lack of accurate information about ICT professional opportunities among high school students when they are making choices about future careers and appropriate Colleges. This situation becomes aggravated because there are not enough certified high school teachers in CS in Western countries, and High School curricula changes do not encourage an emphasis on Math and Sciences.

Often, young people have a misperception of ICT professionals, encouraged by the media and movies (Elias, 2008). They see themselves working in front of a computer, in a cubicle, with little human contact, with long, tedious, and boring shifts (Frieze, 2005)(Lenox & Woratschek, 2008). There's nothing further away from the truth: traditional Engineers and Scientists need large facilities to work, and they are limited by the physical laws that govern our world. A Software Engineer, on the other hand, only needs a computer; all what he / she can imagine can be virtually created inside it. Creativity and innovation are present in every aspect of ICT professionals' work when searching for new solutions. But young people, in general, do not perceive any excitement about ICT, because computer-related subjects in High Schools tend to cover just basic knowledge of office suites (Ali & Shubra, 2010).

In order to bring high school students closer to the professional reality, some universities have developed innovative initiatives, such as those discussed in the following section.

3 New tools in a new era

20 years ago, students taking their first steps in programming were filled with excitement when writing a simple “Hello World” program. Nowadays, our young people are immersed in technology; not only computers, but any type of programmable electronic devices: phones, music players, game consoles, etc. However, in this age of information -in which we have easy access to technology and the skills of design, modeling and programming are essential to the practice of many professions- young people, paradoxically, are becoming more reluctant to explore the possibilities of the machines and to discover how they are made and how to exploit fully their power.

Our environment and the stimuli we receive from it have clearly changed.. Therefore, if we want to raise the student’s curiosity and to train good and motivated professionals, it is necessary to change the approach to reach them. We can take a more attractive path, without losing academic rigor, by using a visual language closer to their way of life and to how they create things.

Several research teams from prestigious Universities have developed learning-oriented visual programming tools for young audiences. These tools have been designed to develop their problem-solving and computational thinking skills by creating interactive animations, games and scientific simulations. It’s worth highlighting the initiatives from MIT Media Lab (Scratch), Colorado University (AgentSheets), University of Kent and Deakin University (Greenfoot), and New York University and Darmuth College (Rapunsel). All these systems have a friendly interface and, in general, their programming languages are basic with a limited repertoire of functions, which facilitates the learning curve in informal education settings.

School students find these tools to be very accessible, and can do many cool things very quickly. But they reach their limits relatively soon, too.

3.1 Rebeca and Alice

"*Rebeca through the looking glass*" is an educational innovation project based on *Alice*. *Alice* and *Rebeca* easily allow to generate 3D animations to tell stories, and to create interactive games and videos that can be later on shared on the Internet, through a friendly programming environment. Both tools have a richer set of capabilities than the previously mentioned software, but students need a lot more initial support and training to fully utilize their object-oriented capabilities.

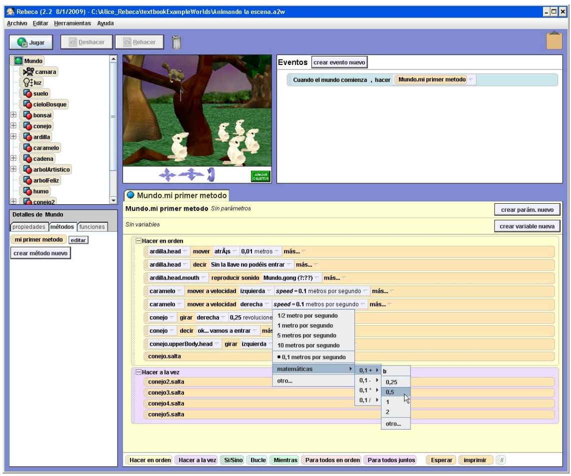


Figure 2: *Rebeca* 3D programming environment.

Alice was initially developed as part of a Virtual Reality research project by the Carnegie Mellon Graphics group, led at that time by Randy Pausch. This project has been evolving over 15 years (Pausch et al., 1995), and it has been supported by numerous Departments, teachers, students and companies.

Alice has been an overwhelming success in the English-speaking community, where this software is used in classes related to Technology by more than a thousand secondary schools (Kelleher et al., 2005, 2007 and 2009).

However, their acceptance in other countries has been rather modest. The reason for this difference is that *Alice* is a software created by and for English speakers. It was not designed to be translated into other languages. Additionally, its structure reflects its 15 years of continuous diverse contributions, becoming a project of more than 170,000 lines of code, largely unstructured in 1900 Java files. For this reason and despite it is one of the most requested features, its creators have dismissed the possibility of supporting other languages.

This is an additional problem for our students, because they not only face the difficulty of learning to program, but they must also deal with a foreign language. Often, the effort to overcome the language barrier is higher than the one needed to assimilate programming concepts.

"*Rebeca through the looking glass*" was born as a challenge in response to this need, with the aim that all young people can benefit from the tool that has risen so many vocations overseas.

3.2 Object-oriented programming learning with Rebeca and Alice

There is an enormous amount of literature about pedagogical studies showing that classical lectures are not as effective as the practice of cooperative techniques and active learning (Ali & Shubra, 2010). The introduction of such practices –promoted by the European Higher Education Area– requires a considerable effort from the teachers, particularly at the beginning, when teaching innovations are incorporated very slowly due to the amount of work involved in assessing student progress.

Learning to program is a hard task for the majority of students, and its complexity is seen as one as one of the main factors that discourages students in the first year (Sloan & Troy, 2008). In general, one of the most enriching and motivating aspects of engineering education is the experimental training that takes place in laboratories, where students can practice their skills and track their progress. However, often object-oriented programming assignments obscure that experience, especially when individual attention in practical classes is not as widespread as it should be (Marris, 2010).

Rebeca and *Alice* provide mechanisms to overcome typical students' difficulties, such us rigid syntax, unfamiliar structure and the amount of time spent to produce a simple output. The drag-and-drop integrated development environment (IDE) of both *Rebeca* and *Alice* eliminates syntax problems that bedevil freshmen. As the student drags-and-drops graphical elements, the source code is constructed and displayed. The student is never permitted to freely edit the code as is the case with most programming development environments. This again shields the student from making syntax errors, but allows them to become familiar with programming language constructs.

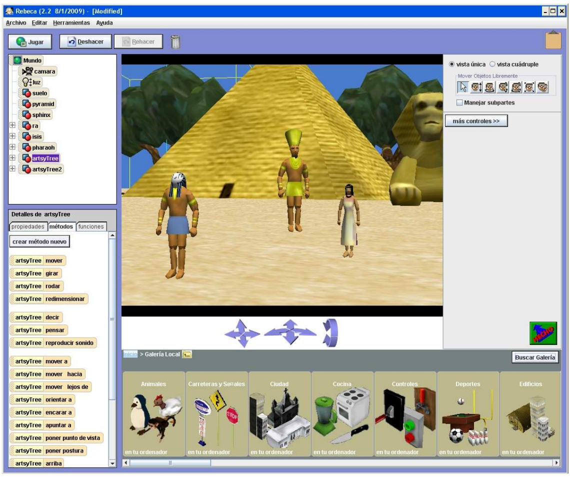


Figure 3: *Rebeca* object gallery.

With *Rebeca* and *Alice*, students can check the status of the program in a much more intuitive way than with a debugger. For example, it is easier to see that an object moves forward rather than backward, than it is to check if a variable has been decremented instead of incremented. Thus, logic errors lead to funny situations that are easily detected. Also, the three-dimensional interface both *Rebeca* and *Alice* present, and the possibilities they offer for creating interactive stories in a collaborative way stimulate the imagination of boys and girls and keep them engaged. In our first pilot courses, we have seen how students do more optional exercises and attend more classes than other groups using traditional tools.

Storytelling is a useful metaphor that can help structuring the process of creating programs in a manner similar to the animation production process. The design is divided in different stages, in which students sequentially develop storyboards, scripts, dialogs, shots and scenes... to turn the story into pseudocode. At the end of this process, students detail their ideas in methods and functions through the interface

Rebeca can be used in High Schools and first year computer programming courses at College. A hybrid approach to teaching *Rebeca* during part of a course and then complementing it by teaching Java during the remainder of the semester may be more effective in reaching educational objectives while helping still to motivate and retain students' interest.

4 *REBECA* DEVELOPMENT

Rebeca has been the result of the joint work of several Graduation Projects offered by the Modeling and Virtual Reality Group at the Rey Juan Carlos University, among which Sergio Ruiz and Irene Montano contributions have been specially relevant. This makes *Rebeca*'s development very special, since it is a piece of software made by and for students with great love and care in details.
Internationalization is the process of designing software so that it can be translated into different languages and regions without the need for further changes to the code. *Alice* has been internationalized and localized as a whole (not just the interface or the tutorials), to allow porting it easily to different regions and languages other than those contained in the original English program.Thus, we had to rewrite thoroughly the original software: more than 1800 files have been modified. As a consequence the localization to any language is now immediate. It only requires the edition of a set of resource files that were added to the original project. In our case the location has been made to Spanish,

but thanks to the prior internationalization process, this software can be ported to any other language.

Since many European languages use characters that do not exist in English, it was necessary to modify the set of applications and how they communicate to use an international character set. Again, this feature has required a laborious task of reverse engineering. Carnegie Mellon distributes *Alice* code under a free open license, but it is not documented and current developers do not give support to make modifications to *Alice* version 2 in official forums.

The development of *Rebeca* has allowed us to fix bugs and add new features, so that the learning experience can be more effective.

A teaching guide has also been written in Spanish (Montano & Ruiz, 2010); this book can be used as an introductory manual for school teachers who want to try *Rebeca* in their courses.

Figure 4: *Rebeca* teaching guide.

5 Deployment in Educational Centers

Several pilot courses, workshops and competitions of different duration have been carried out in a number of events held in Madrid during the last months, such as Juvenalia, fesTICval, and different seminars held at the Rey Juan Carlos University.

The experience has been very positive in all aspects, particularly because of the excellent response of High School students, who stayed in front of their computers even during the breaks. Many of them were amazed to discover that programming is a creative discipline, and they considered a good idea to teach *Rebeca* and object-oriented programming in their schools.

The enthusiasm shown by High School students encouraged us to keep improving this tool. In a near future, we will conduct a study to measure the effectiveness of this tool as an aid in first year programming courses between different groups of students, in order to plan future steps in their development.

6 Conclusions and future work

The result of this project has been a piece of software completely localized to our language, ready to be ported to any other language and ready for use in teaching.

Not only the interface, but also the programming language has been translated. We have designed a programming language similar to Java but with the terminology in Spanish. We have tried to make it as accessible as possible to young audiences, so they can really focus on learning the methodology of object-oriented programming..

"*Rebeca through the looking glass*" and all teaching materials are completely free, and can be downloaded from the official site of the application: **www.gmrv.es/rebeca-es** As soon as possible, the website will also include: a centralized repository where young people can share animations and games made with *Rebeca*, a forum with a think tank who will help solve any questions that may arise in the use of the tool and its use in the courses, and a blog where we will collect suggestions from teachers and students.

ACKNOWLEDGEMENTS

"*Rebeca through the looking glass*" is dedicated to all teachers and students of the Rey Juan Carlos and Carnegie Mellon Universities that made it possible. And especially, to Randy Pausch, so that his legacy can reach all young people worldwide.

REFERENCES

Ali, A., Shubra, C. 2010. Efforts to Reverse the Trend of Enrollment Decline in Computer Science Programs. *The Journal of Issues in Informing Science and Information Technology*, Vol. 7, pages 209 –225.

Bureau. 2010. Bureau of Labor Statistics (2010-11) U.S. Department of Labor, Career Guide to Industries, 2010-11 Edition, "Computer Systems Design and Related Services", retrieved from the Bureau of Labor Statistics web site on June 20, 2010: http://www.bls.gov/oco/cg/cgs033.htm

Dann, W. P., Cooper, S., Paush, R. 2006. *Learning to program with Alice.* Ed Prentice Hall, 1st edition.

Denning, P.J., McGettrick, A. 2005. Recentering computer science. *Commun. ACM*, vol 48 no.11, pages 15–19.

Elias, C. 2008. *La razón estrangulada, la crisis de la ciencia en la sociedad contemporánea*. Ed. Debate. Barcelona, 1st edition.

Frieze, C. 2005. Diversifying the images of computer science: undergraduate women take on the challenge! *SIGCSE Bulletin*, vol 37 no.1, pages 397–400.

Kelleher,C., Pausch, R. 2005. Lowering the barriers to programming: A taxonomy of programming environments and languages for novice programmers. *ACM Computing Surveys*, vol. 37 no. 2, pages 83–137.

Kelleher, C., Pausch, R. 2007. Using Storytelling to Motivate Programming. Communications of the ACM, vol. 50, no. 7, pages 58–64.

Kelleher, C. 2009 Barriers to Programming Engagement. *Advances in Gender and Education*, vol. 1, no. 1, pages 5–10.

Klawe, M., Shneiderman, B. 2005. Crisis and opportunity in computer science. *Commun. ACM*, vol 48 no.11, pages 27–28.

Lenox, T.L. Woratschek, C.R. 2008. Exploring Declining CS/IS/IT Enrollments. Information Systems Education Journal, vol 6, no. 44, 11 pages

Lomerson, W.L., Pollacia, L.F. 2006. Declining CIS Enrollment: An Examination of Pre-College Factors. Information Systems Education Journal, vol 4, no. 35, 13 pages

Montano, I., Ruiz,S. 2010. *Guía didáctica de Rebeca: Aprende a programar con gráficos 3D interactivos*. Ed. Lulu. Madrid. 1st edition.

Marris, E. 2010. US students pay for downturn. Nature, vol. 465, no. 7299, pages 678–679

O'Lander, R. 1996. Factors effecting high school student's choice of computer science as a major. CQL'96: Proceedings of the symposium on Computers and the quality of life. ACM. pages 25–31.

Pausch, R. Burnette, T. Capeheart, A.C. Conway, M. Cosgrove, D. DeLine, R. Durbin, J. Gossweiler, R. Shuichi Koga, White, J. 1995. Alice: Rapid Prototyping System for Virtual Reality, IEEE Computer Graphics and Applications.

Sloan, R. H., Troy, P. 2008. CS 0.5: A better approach to introductory computer science for majors. *Proceedings of the 39th SIGCSE Technical Symposium on Computer Science Education*, pages 271-275.

UNESCO Institute for Statistics. 2010. http://stats.uis.unesco.org/unesco/ReportFolders/ReportFolders.aspx (accessed June, 2010)

Vesgo, J. 2008. Enrollment and degree production at US CS departments drop further in 2006/2007. *CRA Bulletin*.

A MULTI CAMERA SYSTEM FOR AUTOMATIC WARNING DELIVERY IN AMBIENT ASSISTED LIVING APPLICATIONS

Luca Polidori, Ennio Gambi and Susanna Spinsante
Department of Biomedical Engineering, Electronics and Telecommunications, Polytechnic University of Marche, via Brecce Bianche 12, Ancona, Italy
l.polidori@univpm.it, e.gambi@univpm.it, s.spinsante@univpm.it

Keywords: Video Surveillance, Ambient Assisted Living, Multi-Camera System, Background Subtraction.

Abstract: This paper discusses the application of a multi-camera system as a tool for the automatic monitoring of people in indoor environments. More specifically, the target subjects to monitor are those people who, due to ageing or handicap, may experience difficulties in their movements The system should allow the automatic recognition of potentially alarming conditions, such as in the case of falls, in order to enable proper responses, and create a truly assistive environment.

1 INTRODUCTION

Population ageing is a planetary phenomenon confirmed by several authoritative studies, according to which the number of elderly people in the world will dramatically increase in the next future. It is well recognized that Home Automation and Ambient Assisted Living (AAL) solutions, based on the smart environment concept, may provide an effective means to support and facilitate the daily activities of elderly or vulnerable persons. Among the available technologies, advanced computer vision and pattern recognition techniques may contribute to autonomy and well-being of these people, by enhancing safety at home, and maintaining mental and physical stimulation.

In the broad set of ambient intelligent solutions, we focus on indoor localization and tracking of moving subjects, by means of multi-camera systems. Monitoring and surveillance solutions based on computer vision and pattern recognition are developed and implemented since a long time (Cai, 1996), (Yang 2007). In the AAL context, we are interested in the possibility of integrating vision-based systems with other technologies, such as indoor sensor networks, to either reduce the complexity of each processing element, and improve reliability of the outcome provided. This point is particularly important if we consider the final target of any AAL solutions, i.e. to create a safe home environment that is able to automatically react to the user's behaviours, and to anticipate his needs. Surveillance systems shall identify subjects moving in the field of observation, classify them into different groups, and track moving targets during time, according to the who is where is, or the what is where is paradigm. In order to be assistive, these systems should be able to interpret the scene captured, provide a description of the activities performed by the subjects in the scene (Veeraraghavan, 2005), and use such a description to generate events aimed at helping the subject.

In this paper, we discuss a vision based system that tracks the subject and automatically delivers warning messages, when the subject is going to perform potentially dangerous actions. Either the events generating warnings and the way warnings are delivered to the user may be tailored to the subject himself and his needs.

The monitoring system shall work properly under any circumstances and environmental conditions (Hu, 2004), (Kelly, 2009). In the scenario we consider, vision-based applications operate in indoor environments: we can assume an almost quasi-static background, which is a favourable option, as low complexity background-subtraction operations can be reasonably applied. Among the classical problems that are to be faced by vision-based applications (such as colour and light intensity variability, occlusions, rapidly moving subjects, differentiated sights and shots of the same object, movements of the subjects in the scene), segmentation and tracking of several subjects located in the same scene at the same time, with possible occlusions occurring, represent a critical issue. Implementing a multi-camera solution may solve most of the issues raised by such conditions, even if complexity and required processing resources increase.

This paper presents the application and the performance obtainable by a multi-camera computer vision solution in the context of AAL, aimed at monitoring the behaviour of elderly or partially impaired people in indoor environments (such as their own home), and analysing their attitudes, in order to recognize possible suspicious situations denoting health problems, danger, or safety risks and to generate multimodal warnings.

By this way, it is expected to support assisted living applications that can automatically react to any potential alarming event, such as in the case the subject has not kept moving for a long time, or he is recognized, through the computer vision system, as lying on the ground. As initially stated, the functionalities foreseen by the proposed system are difficult to implement when no constraints are imposed on system itself, on the environment, and on the subjects to monitor. In the proposed application, however, it is possible to set some initial hypotheses that are helpful to limit the complexity of each processing step. In details, we assume:

a) a multi-camera system with a wide overlapping area;

b) an indoor environment, made up of rooms with rectangular shape, and a quasi-static background;

c) the only subjects to monitor are people that are assumed to be in motion, with a random behaviour.

The final target of the application, under the above hypotheses, is to identify a moving subject in the scene, to obtain his/her planar position, and to track the temporal evolution of the planar position .Understanding the planar position of the subject, his movements, and the direction followed, allows to generate possible alarms in the case of dangerous directions or actions. Further, it is important to identify subjects that are not moving in the scene, and to evaluate their eventually suspicious positions due, for example, to falls on the ground.

The outcomes provided by the automatic monitoring system should be valid even in the case of multiple subjects in the scene, i.e. when possible occlusions may occur. The main steps of the proposed approach are graphically represented in Fig. 1, and detailed in the following sections.

The paper is organized as follows: Section 2 deals with the issue of foreground segmentation and related problems; Section 3 discusses the projection of the multi-camera system onto the ground plane, in order to locate the position of the subject, and track him, by means of proper techniques presented in Section 4. Finally, Section 5 concludes the paper.

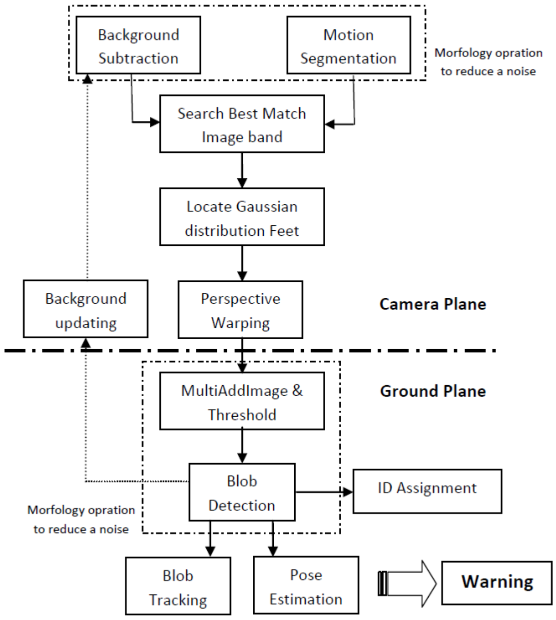


Figure 1: Main processing steps of the proposed approach

2 FOREGROUND SEGMENTATION

The proposed system aims at monitoring subjects that shall be located on the visual scene by means of a background subtraction process (Moeslund, 2000), or a motion-based segmentation (Lipton, 1998), (Yang, 2005). The algorithms applied are automatically chosen according to the specific scenario to analyze. In some cases, a combination of the different solutions may provide the best performances.

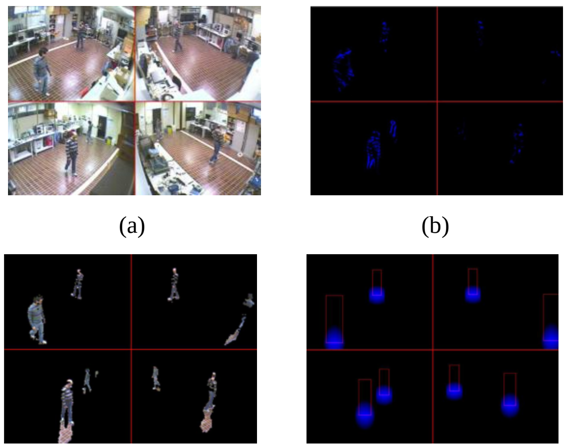
In our context of interest, the targets are all those elements that can move independently within the observed environment. However, due to the fact that we are focusing on an assisted environment, we shall be able to localize subjects even when they do not move within the scene, by tracing their location and evaluating the nature of the position they assume. Moving subjects are located by means of image segmentation, that can be achieved through a preliminary representation of the scene, named background model; the background model is then compared to each single frame of the video sequence captured during time, in order to highlight the differences among them. Any variations in the frame, with respect to the background model, denote the presence of a target (subject) of interest. Pixels belonging to that area define a shape, named blob, i.e. a specific part of the frame, also called foreground.

Limiting our analysis to an indoor environment, it is possible to add further constraints on the system. We can assume a static frame, clear of subjects, as our background model, or, at least, a quasi-static frame, where some elements may show periodic variations, but result static during the observation time (e.g. light variations). In the software implementation designed for simulation, the background model is a weighed time average of previous frames, the number of which depends on the frame rate set by the capturing device. The foreground is obtained by applying a motion detection algorithm that subtracts the background model to the current frame, and collects the varying pixels as the shape looked for. Foreground segmentation performed through motion detection may suffer a number of limitations, that reduce the process precision. They are mainly due: a) to not evident motion, such as in the case of subjects moving towards the camera; b) to the presence of many subjects in the scene, that may cause occlusions; c) to the colour similarity between the subjects' clothes and the background.

Shadows and light reflections on the floor may also reduce the precision of the segmentation process, especially in the case of indoor environments, like the one we are focusing on. In order to improve the efficiency of our global implementation, we did not focus on the problem of obtaining well-defined shapes from the foreground segmentation step, but rather on the idea of locating those areas of the frame where subjects are present, to estimate the related probability of occupation. Fig. 2 a) shows four different views of the same scene, obtained by four different cameras. Figs, 2b) and 2c) show, respectively, the results provided by the motion detection and background subtraction algorithms, and the critical features previously discussed.

Our approach, aimed at reducing complexity and required resources, is able to provide correct localization of the subjects in the scene, even when the segmentation outputs are less precise, as shown in Fig 2 d).

Rectangular shapes, of different dimensions, have been adopted to represent each subject located in the frame by the segmentation process. They denote the probability a subject is present in a given area of the image. By this solution, we can get a clear and simple result when moving to the virtual top view plane, obtained by means of the projection onto the ground plane. Moreover, as it will be discussed at the end of this Section, the dimensions of the rectangular shapes are tuned to the position of the segmented subject in the scene, i.e. nearer or farther to the camera, and even weighted by means of a Gaussian profile, to better evidence the position of the subject's feet on the ground plane. This peculiar expedient is very useful when dealing with the overlapping of the frame projections, generated by different cameras, on the ground plane. In the specific context under evaluation, i.e. indoor and home environments, we can limit our interest to the ground plane, instead of the usually adopted head plane. In fact, we can reasonably assume that the visual scene is not crowded (so that the subjects' feet could not be revealed), and, most of all, we are interested in the possibility of automatically recognizing subjects lying on the floor, which motivates our focus on the ground plane projection.



(a) (b)

(c) (d)

Figure 2: (a) Source Image; (b) motion segmentation; (c) background subtraction; (d) Feet Detection Probability

The global foreground segmentation process, based on background subtraction and motion detection, implemented in our software application (Fig2 b) and c)), comprises the following steps:

- each frame processed by the motion detection algorithm is subjected to filtering, in order to reduce the impact of noise. Among the filtering operations, morphology algorithms are iteratively applied to improve the image quality. A threshold is also set, to discriminate valid pixels from non-valid ones;

- the average value of each column of pixels in the frame is computed, by associating greater weight to the top pixels and smaller weigh to the bottom ones. This is motivated by the fact that top pixels are related to shapes more distant from the camera, so smaller than shapes located nearer to the camera. The average values are stored in a numeric array, whose length equals the number of columns in the frame. The average column values are computed as follows:

$$v_{avg} = k \cdot v_{tot} \cdot (p_f - p_t) \qquad (1)$$

$$v_{tot} = \frac{\sum_{i=0}^{H} (H - i) \cdot p_i}{H} \qquad (2)$$

where: p_f and p_t denote the first and the last pixel of value different from 0 in the column, from top to bottom; H is the total number of pixels in the column; the index i moves along the column, from top (i = 0) to bottom (i = H); k is a factor due to normalization;

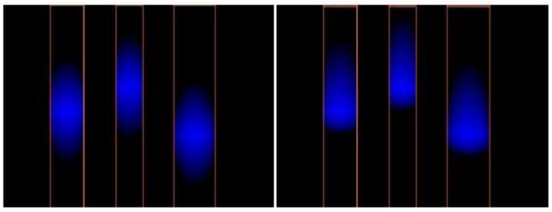
- a sliding window of fixed width is applied to the numeric array; a new array of the same length is obtained, where each element is the average of the values in the first array, falling into the window;
- a threshold is applied to the new array, and the values lower than the threshold are set to zero. All the pixels in the same frame column corresponding to each element of the array are set to the value of the element itself;
- the relative maximum values within the numeric array are located, and the corresponding column indexes saved. They correspond to the shapes in the frame we are looking for.

At the end of the process, the number of relative maximum values found within the array should equal the number of shapes in the frame. If two (or more) shapes are very close to each other, or partially overlapped, a single relative maximum is found instead of two (or more than one), and the shapes may also become not separable.

A further feature we added to the segmentation process to limit, and even avoid, the presence of false positives in the top view plane, consists in varying the dimensions of the rectangular shapes to account for the distance of the subject from the camera. The final dimensions of the rectangular shape depend from the location of the subject's feet pixels. The shape width is chosen equal to 1/3 of the height.

Due to the non-zero probability of light shadows or reflections from the ground (such as shown in Fig. 2c)), a useful work around for the correct location of the subject's feet is to locate the head of the subject within the band, and then to apply a reasonable proportion with respect to the width and height of the shape (i.e. his distance) in the image. In fact, the position of the head is almost free from the disturbances discussed, and its localization may provide more precise results.

As previously stated, in our approach each shape denotes a probability of presence of the subject in the frame area. In order to improve the precision of the location process, performed on the ground plane projection, rectangular shapes are shaded off, by filtering the pixel values from the bottom to the top, and from the centre to the side edges. Filtering may be performed according to a Gaussian profile, as shown in Fig. 3 a), or according to different profiles (see an example of "Drop-like" profile in Fig. 3 b)); the solution chosen is the one which maximizes the precision of the localization step on the ground plane projection. The vanishing effects depend on the position of the shapes in the image: the standard deviation which defines the vanishing of the shapes, along either the x and y directions, is determined by the latitude of the shape base point, i.e. by its distance from the camera.



a) b)

Figure 3: Filtering profiles: a) Gaussian profile, b) "Drop-like" profile.

3 PROJECTION ONTO GROUND PLANE

The use of a multi-camera system in the target indoor environment is motivated by the need of avoiding possible occlusions, due to overlapping foregrounds. As a matter of fact, occlusions are avoided by mounting the video capture device above the moving subjects: the dimensions of the subjects, and the crowded zones within the monitored environment, may be easily estimated. However, video capture devices positioned above have a reduced field of vision, so that many capture devices must be used to cover the whole environment to monitor. The advantages obtainable by a camera placed above motivate the idea of projecting the captured image as if taken by a virtual camera, in a so-called top view (Taj, 2009), (Mittal, 2001). A top view shot is obtained as the projection onto the ground plane (referred to as the z = 0 plane) of the frame captured by a video camera.

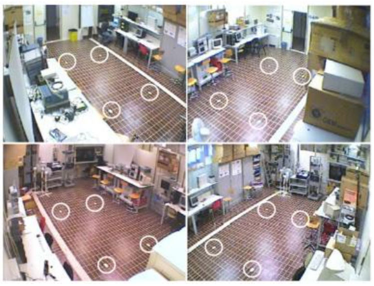


Figure 4: Reference points extracted from different views of the same scene.

The final image dimensions, related to the virtual camera elevation that affects, in its turn, the measurement resolution and the extension of the global field of view, are arbitrary, and mathematically expressed by the scaling factor obtainable by the projection matrix, named homography matrix. An homographic transform allows projecting a plane into a corresponding perspective image. The associated homography matrix provides the position, on the destination image, of the points selected in the original image. In our specific application, we need to locate some points extracted from the images captured by different cameras on the ground plane, as shown in Fig. 4: consequently, we need to find the related homography matrices, that can be obtained from the different cameras' views of the same scene. In a general model, an homography transform is described by a non-singular 3 x 3 matrix H:

$$\lambda \cdot \begin{bmatrix} x_1' \\ x_2' \\ 1 \end{bmatrix} = \begin{bmatrix} H_{1,1} & H_{1,2} & H_{1,3} \\ H_{2,1} & H_{2,2} & H_{2,3} \\ H_{3,1} & H_{3,2} & H_{3,3} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} \qquad (3)$$

where λ is a scaling factor, and matrix H has 8 degrees of freedom. Points are described by homogeneous coordinates (x1; x2; x3), where (x1/x3; x2/x3) are the corresponding Cartesian coordinates.

In the proposed application, we compute H without any information about the intrinsic parameters of the cameras, i.e. calibration information, thanks to the availability of multiple views. We implemented a software function, named:

```
funcFindHomography ( src_points,
    dst_points, homography ),
```

that collects two sets of corresponding points, in homogeneous coordinates, from two different views of the same image, and retrieves the best homography that minimizes the backprojection error in the correspondence. To apply the function, at least four couples of points are needed; anyway, the higher the number of available couples, the better the homography found works. Once the homography matrix between each camera view and the common ground plane has been computed, it is possible to project each image point (i.e. each pixel) into the virtual top view.

This operation is performed by means of a second software function, named:

```
funcWarpPerspective ( src_image,
    dst_image, homography ),
```

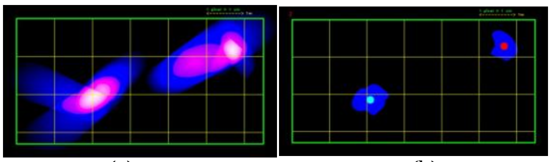
to obtain the perspective transform. The function gets in input the captured frame and the homography matrix, and outputs the same frame projected on the ground plane. Given the perspective view of the scene, the resulting image will be distorted, and its pixels obtained through an interpolation process that may require different time, according to the accuracy of the operation. All the views captured by the different cameras shall be used to obtain the full frame projection onto the ground plane.

Figure 5: Warping of different views of the same scene, and ground plane projection.

4 LOCALIZATION AND TRACKING

The projection onto the ground plane of the different views captured by the multi-camera system evidences a common area generated by the overlap of the projected views. When the warping operation described by the homography matrix is applied to the video frames captured by the different cameras, their joint projection onto the ground plane is obtained, as shown in Fig. 5.

Our idea is to apply the warping transformation not to each whole frame, but only to the foreground images obtained from each frame, by the segmentation process. By this way, the overlap, intersection and merging of the projected foreground images, identify a common area in the virtual top view plane, that corresponds to the planar position of the subject moving in the environment, i.e. to the lower part of his body (legs and feet). More precisely, the common area is related to the probability of occupation of a given zone, that increases if the number of blobs contributing to the overlap increases. An example is shown in Fig. 6 a): the magenta and white spots in the image represent the areas where the probability of presence of a subject is higher. Light blue areas denote possible “false positives” that are easily avoided, by means of thresholding and morphology operations to reduce noise contribution and filter image, as shown in Fig. 6 b).



(a) (b)

Figure 6: Subjects’ localization on the projected ground plane: a) Projection output; b) Quantitative localization after thresholding and filtering.

We are interested in localizing the subject, and tracking his movements, so we focus our processing on the overlapping area in the ground plane projection. Quantitative measurements of the positions occupied by the moving subject, and the distances covered, may be obtained by exploiting the homography matrices, and the known dimensions of a reference element in the image, such as a carpet, or a tile; alternatively, the calibration parameters of the whole multi-camera system should be available. In our implementation, we exploit the information about the dimensions of some reference objects to draw the subject’s coordinates on the ground plane, and track his trajectories.

The final aim of our application is to generate alarms when dangerous conditions are revealed. They may be determined by either suspect directions of movements, or unnatural positions assumed by the subject on the ground (e.g. denoting a fall). It is consequently necessary to track the subject’s movements, and store in a state structure the positions assumed during the observation period.

To track the position of the monitored subject on the planar view, i.e. on the ground plane projection obtained by means of the processing steps previously discussed, we can apply a tracking algorithm directly on the ground plane projection. This solution reduces the complexity of the process, because the tracking algorithm shall be applied on a single image only, and not on a set of corresponding images. Further, blobs projected onto the ground plane do not vary significantly, and can be tracked by means of rigid movements.

The solution chosen for tracking purposes, at least in this preliminary phase of the study, is a well established and well known algorithm, the details of which may be found in (Bin, 2009). Kalman filtering (Welch, 1995) is also used, in order to track the current subject’s position, and even foresee his/her future movements inside the monitored environment.

5 CONCLUSION

This paper presented a multi-camera solution for the automatic monitoring of people in indoor environments, aimed at supporting the implementation of assistive services, through a low complexity software processing, and a low cost hardware.

The proposed solution relies on a simplified image segmentation step, by means of which it is possible to locate the presence of a subject in the scene with the highest probability; then, the following tracking operations may be performed on a single view, obtained through a projection onto a virtual ground plane.

By analyzing the virtual ground plane only, we can in fact distinguish the objects included in the scene and identify them, to determine their instantaneous position, their movements along the sequence of frames, their future trajectory, and possible suspicious behaviors, in order to generate corresponding alarms. By univocally associating the virtual ground plane and the views captured by each

single camera, it is possible to know exactly at each moment how many people are present in the room, and what areas of the room are affected by their presence. Storing the number and the states of the subjects located in the room and revealing if they are standing or laying on the ground allow to properly generate and issue specific alarms.
Further research activities are currently ongoing, to test the effectiveness of the scheme, and compare its performance to those obtained by similar solutions proposed in the technical literature.

ACKNOWLEDGEMENTS

The authors would like to thank Antonio Petito for his contribution to the development of the software application adopted in the project herein presented.

REFERENCES

Cai, Q., Aggarwal, J.K., 1996. Tracking human motion using multiple cameras, *International Conference on Pattern Recognition*, vol. 3, pp. 68 - 72, Vienna, (Austria).

Yang , T., Chen, F., Kimber, D., Vaughan, J., 2007, Robust People Detection And Tracking In A Multi-Camera Indoor Visual Surveillance System, *Proc. ICME 2007, IEEE International Conference on Multimedia And Expo*, Bejing (China).

Veeraraghavan, A., Roy-Chowdhury, A. K., Chellappa R., 2005, Matching shape sequences in video with applications in human movement analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1896 1909.

Hu, W., Tan, T., Wang, L., Maybank S., 2004, A survey on visual surveillance of object motion and behaviours, *IEEE Transactions on Systems, Man, and Cybernetics*, Part C, vol. 34, pp. 334 - 352.

Kelly, P., O’Connor, N. E., Smeaton, A. F., 2009, Robust Pedestrian Detection and Tracking in Crowded Scene, *Image and Vision Computing Journal*, pp. 1445 - 1458.

Moeslund , T. B., Granum, E., 2000, Multiple cues used in model-based human motion capture, *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 362 - 367, Grenoble, France.

Lipton, A. J., Fujiyoshi, H., Patil, R. S., 1998, Moving target classification and tracking from real-time video, *IEEE Workshop on Applications of Computer Vision*, pp. 8 - 14, Princeton, New Jersey.

Yang, M-T., Shih, Y-C., Wang, S-C., 2004, People tracking by integrating multiple features, *International Conference on Pattern Recognition*, volume 4, pp. 929 - 932, Cambridge, UK.

Taj, M., Cavallaro, A., 2009, Multi-camera track-before-detect, *ACM/IEEE Int. Conf. on Distributed Smart Cameras (ICDSC)*, Como, Italy.

Mittal, A., Davis, L., 2001, Unified Multi-camera Detection and Tracking Using Region-Matching, *IEEE Workshop on Multi-Object Tracking (WOMOT’01)*, Vancouver, Canada.

Bin, Z., Jun-Zheng, W., Jing, L., Wei, S., 2009, Object Tracking with Mean Shift and Model Prediction, *International Congress on Image and Signal Processing*, 17 - 19, pp. 1 - 5.

Welch, G., Bishop, G., 1995, An Introduction to the Kalman Filter, *Technical Report: TR95-041*, University of North Carolina at Chapel Hill, USA,

MONITORING ACCESSIBILITY IN DIGITAL TELEVISION
Subtitling, audiodescription and EPG

Mercedes de Castro, Belén Ruiz
Carlos III University of Madrid and Spanish Centre of Subtitling and Audiodescription, Avda.Universidad 30, Leganés, Madrid, Spain
mcastro@it.uc3m.es, bruiz@inf.uc3m.es

Carlos Alberto Martín, Lara García, José Manuel Menéndez
Universidad Politécnica de Madrid, Ciudad Universitaria s/n, Madrid, Spain
cam@gatv.ssr.upm.es,lgv@gatv.ssr.upm.es,jmm@gatv.ssr.upm.es

Keywords: Digital TV, IPTV, Subtitling, Accessibility, EPG, Monitoring

Abstract: This paper presents methodologies and tools developed to monitor the accessibility services in digital television in a moment when the principles of design for all are being considered in all new audiovisual media communication services. The recent approval in Spain of the General Communication Audiovisual Law (LGCA) establishes the need for a monitoring activity that ensures fulfilment of minimum level of accessibility services, thus giving special relevance to this research. The main objective of Hermes-TDT project is to measure the quality and quantity of accessibility services offered by digital television delivered, its first version focusing on digital terrestrial television. Preliminary results, presented here, consist of a DTT signal sniffer prototype and the methodology for quantity and quality measurements. The results can be useful to regulatory entities and broadcasters to measure both quantity and quality of the accessibility services offered to the users, helping them to set the priorities to improve services currently available to users.

1 INTRODUCTION

Accessibility services as a way to access the audio or video content of multimedia are required for an increasing number of people with hearing or visual impairments, people that need some help to understand the language or people experiencing multimedia in noisy environments or places where the audio must be turned off.

The need to provide accessibility services as a necessary complement to all new technologies has been recognised by the UN (United Nations 2006; Kayess et al. 2008), and governments in many countries have taken initiatives to promote and regulate accessibility services to multimedia television contents (Ofcom 2008; OSI/EU 2005). This is the case of Spain where the recent approval of the General Communication Audiovisual Law, LGCA, (Spanish Government 2010) establishes minimum levels of availability of accessible multimedia in DTT and also in the fast growing IPTV networks.

Once regulations are established, the need to measure and qualify those services is a natural consequence and a requirement for all involved actors. Television content providers need to evaluate the real presence of accessibility in their offers, users' organizations are also interested parties, but, for regulators, reliable monitoring is a responsibility.

To measure the degree of compliance with obligations that, related to the requirements of the users, are imposed on television operators, tools and research methodologies are necessary to monitor both the quantity and quality of the provided services. This challenge has been addressed within the Hermes-TDT project of which preliminary results are presented in this article. The methodology and technology developed in Hermes-TDT allow recognizing and characterizing the content of the accessibility services in digital television broadcast (terrestrial, satellite...) and can be extended to IPTV television networks. This is something that is in the interest of television content providers, users of accessibility services, regulators of the audiovisual sector and related research communities. This project will also help to define new courses of action in digital television to improve the existing accessibility services and to implement new ones.

This article describes, within the context above, methodologies and tools, which developed within the Hermes-TDT project, are applied to monitor subtitling and audio description services in DTT and can be applied also to IPTV. Hermes-TDT not only provides an accurate monitoring of the quantity of offered services but also captures data that can be used to define and measure improvements in quality.

2 MONITORING PARAMETERS

Accessibility to digital television is provided, as a first step, at the emission side by the content providers. Digital television accessibility services are mainly subtitling, audiodescription and sign language.

The European standard for Digital Television (DVB) contemplates two types of subtitles: DVB subtitling and Teletext (Tanton & Weitzel 1999; Martin Edo et al. 2007; Clark 2004). Some European broadcasters such as BBC or CRTVE use simultaneously both technologies while other broadcasters have chosen one of the two. In both cases, subtitling signalling in the Transport Stream (TS) of the television signal is an important parameter under analysis, needed to verify the interoperability and to ensure the technical quality of the reception. Subtitling signalling is part of the PSI and SI tables of the DVB/MPEG-2 Transport Stream and its presence and actual values may not be fully in line with the actual content of the subtitle streams themselves. A typical television program screenshot with the subtitles for the hard-of-hearing turned on is shown in Figure 1.

Figure 1: Subtitles for the hard of hearing

The audio description provides a narrative of the visual elements of an audiovisual program to visually impaired people (Utray et al. 2009). The approach used in Europe to deliver this accessibility service is the use of an alternate audio channel to provide a mix of the original soundtrack and the audio description. In the United Kingdom, a system of local mixing in the receiver, also recognized by the DVB digital television standard (Stallard, 2003: 22), is available as an alternative. A relevant aspect in this research is the analysis of audio description signalling within the Transport Stream. Signalling, defined in the standards (ETSI 2003; ISO/IEC 2007), is not homogeneously used in the different countries of the European Union. There are even different interpretations of the standard within the countries themselves.

For some hearing impaired people, sign language is a basic accessibility aid to access the audio content of a television program. Although combinations of sign language and different technologies have been widely studied (Pérez-Ugena 2008; Baez & Fernández 2010; Pereira 2010), the use of sign language in television is currently limited to the incorporation of an interpreter in the image, either integrating her/him in the original television staging or using a dedicated window in the screen.

A closed sign language option that users might enable or disable using the remote control is a digital television functionality still pending of development and integration. This fact does restrict the automatic monitoring of this service because, at present, sign language interpretation services are embedded (open sign language) in the video signal and therefore not identified as such in the transmission stream.

2.1 Parameter definition

Within the framework of the different accessibility services, this research focuses on the provision of subtitling and audio description by television content providers.

Tools and procedures for measuring and typifying these services, including the definition of monitoring parameters, must consider both the aspects that can be automatically measured and those that can only be measured with human intervention.

The MPEG-2 (ISO/IEC 2007) structure consists of data streams (video, audio, data) and signalling streams and, of these, some are PSI signalling tables (PAT, PMT, …) while other, in the case of Transport Streams carrying DVB digital television (Reimers 2006; Stienstra 2006), are SI tables (EIT, SDT,…). The monitorization, be it automatically or human based, requires scanning several sources of information:

- Transport Stream composition as defined in PSI tables (ISO/IEC 2007)
- Transport Stream actual content of elementary streams
- DVB program and service information tables (SI): EIT, SDT...(ETSI 2003)

In the group of parameters that can be automatically measured, we consider the existence of an audio stream for audiodescription, existence of the subtitle streams, actual presence of subtitles in the television channels, EPG content, coherence between subtitles presence and signalling data, colour use in subtitles and subtitle speed. And, specially, numbers of hours of subtitled and audiodescribed programs.

Table 1 shows a summary of the main parameters that can be automatically measured in the digital television signal:

Table 1: Variables to measure accessibility automatically

Variable	Evaluation
Subtitle presence/absence	Based on actual contents of subtitle elementary streams
Audiodescription presence/absence	Based on EIT, SDT and PMT descriptors
Total subtitling time including spots	Compute the total time of subtitle actual presence in the Transport Stream
Total audiodescription time including spots	Compute the total time of signalled audio containing audiodescription based on EIT, SDT and PMT descriptors
Historic EPG	Actual Transport Stream, EIT present&following
Coherence between EPG and actual subtitles	Cross-checking between actual content of subtitle elementary streams and SI and PSI tables
Character per line in the subtitles	Number of lines and characters per line on screen
Subtitle speed	Words per minute in subtitled programs
Subtitle colors	Colour detection in Teletext and/or DVB-Sub subtitles
Subtitle text	Subtitle text capture from Teletext and/or OCR from subtitle images in DVB-Sub elementary streams

In the group of parameters that must be measured with human intervention we consider the advertisement detection in the video streams (that in some scenarios could be automated), actual content of the audiodescription streams, literality, and subtitling of audio effects…see Table 2.

Table 2: Variables to measure accessibility with human intervention

Variable	Evaluation
Program name in EPG	Check if name in EPG corresponds with program name
EPG synchronization	Check if EPG information is updated when a program starts, within a predetermined time window.
Subtitling signalling	Check that subtitle icons appear in the On Screen Display of the television set
Presence of subtitles	Activate Teletext subtitles and check they appear on screen
	Activate DVB subtitles and check they appear on screen
Subtitle position on screen	Check that speech subtitles are centred and in the lower part of the screen
	Check position of sound effects subtitles
	Check that no overlays occur when also open

	captions appear on screen
No. of lines of text in a subtitle	Get the maximum number of lines simultaneously displayed on screen in pre-recorded programs
	Get the maximum number of lines simultaneously displayed on screen in live programs
Colour usage	Check whether colours are used to identify different characters in the program
	Check if colour allocation is consistent for the whole program
Subtitling speed	Check the time individual subtitles are displayed and their length to obtain maximum and average subtitling speed
Subtitle grammatical correctness	Evaluate the level of orthographical and syntactical correctness of text
Subtitle literalness	Evaluate how close subtitle and speech content are
Subtitle synchronization	Measure subtitle presentation times relative to corresponding speech fragments to obtain maximum an average offsets in pre-recorded programs
	Measure subtitle presentation times relative to corresponding speech fragments to obtain maximum and average offsets in live programs

By measuring those parameters it is possible to typify the quantity and quality of accesibility services in digital television.

3 HERMES-TDT SYSTEM

A system has been developed called 'DTT signal sniffer' that automatically analyses the signal transmitted by broadcasters extracting data and signaling related to the services mentioned above. This system, in a prototype phase, allows quantitative measurements of the services and also verifies the adequacy of signalling with regard to the applicable television broadcasting quality standards. Hermes –TDT monitoring architecture supports two basic monitoring steps:

a) Real time Data Acquisition, and

b) Storage and Data Analysis.

The first step is carried out by the Hermes TDT Signal Sniffer, a distributed system able to continuously capture the relevant information from DTT signal in various geographic positions. The information is then transferred and stored in a central database. Each Hermes-TDT signal sniffer node explores the DTT signal. New information is time-stamped and stored in the database. Relevant stored data are: DVB subtitles, Teletext subtitles, EPG information, television channel composition. Optionally, complete program including video, audio, audio description and subtitles can be recorded. All the information is extracted from the Transport Streams

The system is depicted in Figure 2:

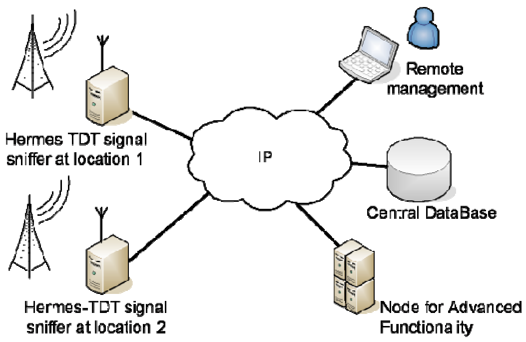


Figure 2: Hermes-TDT signal sniffer

During the analysis phase the stored information can be post-processed to obtain measurements on: Subtitle presence/absence, Audio description presence/absence, total subtitling time, coherence between EPG and actual subtitles, subtitle colours, speed, length, historic EPG. A web application running in the central node allows accessing to stored data from a remote host, as well as the management functions of such distributed system. The main screen to access these functionalities is shown in Figure 3 below:

Figure 3: Hermes-TDT navigation menu for Historic EPG and subtitling queries on stored data

Also human-based analysis are possible based on the stored data: Speech and subtitle synchronization, subtitle colour for the identification of speakers, literalness of subtitles, verification of adequacy for the deaf and hard of hearing, actual audio description presence in the alternate audio streams, etc.

4 QUANTITY OF SUBTITLING

Quantity of subtitling is among all aspects related to accessibility to television the most relevant from the users's point of view and for regulators. In this research, a procedure has been designed to measure the quantity of time that subtitling is available in IPTV or DTT channels. For this reason, the first objective considered is to verify if broadcasters are complying with the percentage of accessible programming defined by the regulator. In Spain, the level of service is defined in the General Communication Audiovisual Law (Spanish Government 2010).

When dealing with quantity measurements of subtitling time in DTT broadcast or IPTV networks, the MPEG-2 transport stream composition and content, including signalling (ETSI 2009), need to be explored in detail.. Subtitle data are carried as part of the television channel, as a private data elementary stream within the MPEG-2 streams (ETSI 2005). DVB foresees two ways to transmit subtitles: in Teletext (ETSI 2003) or digital format (ETSI 2009). Taking into account the standards involved, three ways to measure the actual emission of subtitling exist:

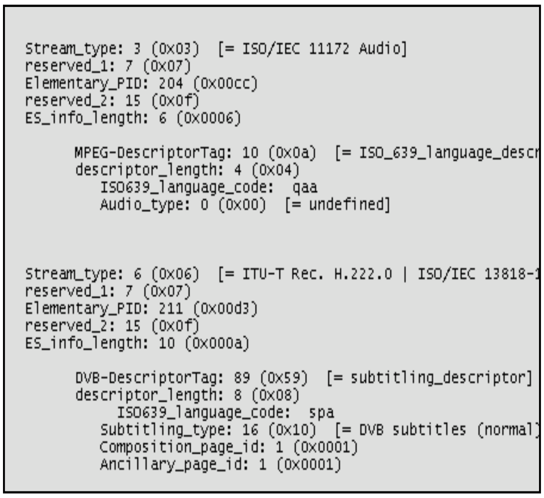
a) Exploring the PSI tables: look for the existence of an elementary stream for subtitles in the same PMT that defines a TV channel in the Transport Stream; subtitles can be DVB-Sub and/or Teletext subtitles.

b) Exploring the SI tables: the indication of subtitle availability and subtitle type in the SDT or EIT tables

c) Exploring the stream content: monitoring the actual subtitle presence in the corresponding data elementary streams carrying the subtitles

Procedures a) and b) require that all PSI or SI tables transmitted be aligned with the actual content of the transport stream. This is not always the case in the deployment scenarios of DTT television. In Spain, one of the EU countries that has recently completed the analogue TV switch-off, the contents of the PSI and SI tables are not fully consistent with the actual content in the video, audio and subtitle streams. This is one of the preliminary results of the research done in this project.

Method a) consists of computing the total time that elementary streams for subtitling are present in the MPEG-2/DVB signal in a period of time. It is based on the fact that TV channel composition (formalized in its PMT) should vary in real time to match the actual multimedia content delivered in that channel. If such practice were fully implemented, PMT tables would reflect the presence or absence of subtitles at program level by adding or removing the references to the PIDs of the corresponding DVB-Sub or Teletext elementary streams (Forbes, 2001). However, it is unusual that broadcasters remove from the PMT table of a TV channel the reference to the subtitle streams when these are not carrying subtitles. Although subtitle unavailability is frequent (e.g. in live programs, ads, sports programs...), it is normal practice to maintain the references to these temporary empty subtitle streams in the PMT tables. This mismatch between program composition in PMT and subtitle data actual content appears in Figure 4 and Figure 5 that show two snapshots of a capture from the DTT broadcast of a TV channel in Spain. Figure 4 shows the contents of the PMT tables of channel “La 2”, pointing to the subtitle stream.



```
Stream_type: 3 (0x03)  [= ISO/IEC 11172 Audio]
reserved_1: 7 (0x07)
Elementary_PID: 204 (0x00cc)
reserved_2: 15 (0x0f)
ES_info_length: 6 (0x0006)

      MPEG-DescriptorTag: 10 (0x0a)  [= ISO_639_language_descr
      descriptor_length: 4 (0x04)
         ISO639_language_code:  qaa
         Audio_type: 0 (0x00)  [= undefined]


Stream_type: 6 (0x06)  [= ITU-T Rec. H.222.0 | ISO/IEC 13818-1
reserved_1: 7 (0x07)
Elementary_PID: 211 (0x00d3)
reserved_2: 15 (0x0f)
ES_info_length: 10 (0x000a)

      DVB-DescriptorTag: 89 (0x59)  [= subtitling_descriptor]
      descriptor_length: 8 (0x08)
           ISO639_language_code:  spa
         Subtitling_type: 16 (0x10)  [= DVB subtitles (normal)
         Composition_page_id: 1 (0x0001)
         Ancillary_page_id: 1 (0x0001)
```

Figure 4 - PMT table for TV channel “La 2” from RTVE in Spain (10 March 2010)

Figure 5 shows the bit rate of the subtitle stream for channel “La 1” (green area), with actual subtitles being broadcast, and “La 2” (blue area), where no subtitles were available. It can be observed that “La 2”, although having a dedicated stream for subtitles which carries residual information, does not convey actual subtitle data.

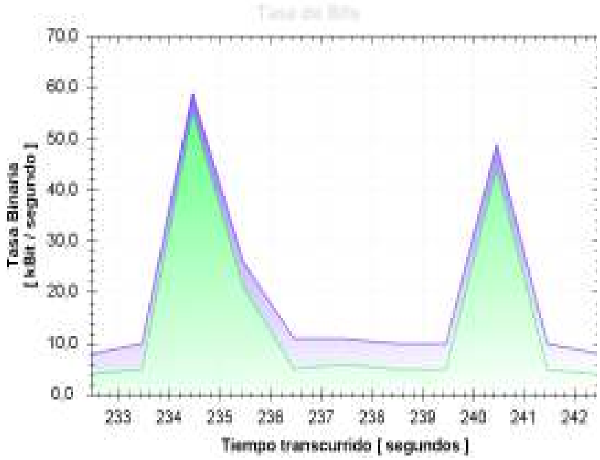


Figure 5 - Bit rate of DVB-Sub stream for TV channel “La 1” (green) and “La 2” (blue)

It can be then concluded that, in the present implementation stage, method a) is not appropriate to obtain quantity measurements in TV emissions. An alternative to method a) is method b), with measurements based on Service Information (SI) tables of the Digital Video Broadcasting Project (DVB). Among the data structures and descriptors in the SI tables (ETSI 2009), DVB foresees a *component_descriptor* that can be added to the EIT or SDT tables to signal the presence and type of subtitles in a TV program. It identifies the type of multimedia content carried over an elementary stream of a TV channel, and may be used in the broadcast EIT table to provide a text description of the elementary stream content. Table 1 shows values of *component_descriptor* relevant to the description of subtitle elementary streams:

Table 1: Subset of the values of fields in the component_descriptor of the DVB SI

Stream type	Component type	Description
0x03	0x01	EBU Teletext subtitles
0x03	0x02	associated EBU Teletext
0x03	0x10	DVB subtitles (normal) with no monitor aspect ratio criticality
0x03	0x14	DVB subtitles (normal) for display on a high definition monitor
0x03	0x20	DVB subtitles (for the hard of hearing) with no monitor aspect ratio criticality
0x03	0x24	DVB subtitles (for the hard of hearing) for display on a high definition monitor

DTT or IPTV receivers use the *component_descriptor* in the EPG menu and On Screen Displays (OSD), in the user navigation menus, and, in some cases, to extract and handle the data in the elementary streams. When properly signalled, the existence and details contained in the *component_descriptor* indicate the existence and type of subtitles in a TV program. As in case a), this is not yet normal practice in the DTT broadcast in Spain. In fact no national TV broadcaster in Spain signals the existence of the subtitle elementary streams by means of the *component_descriptor* neither in the EIT tables nor in the SDT.

Having explored alternatives a) and b) in detail, the conclusion is that measuring net subtitling time in the TV broadcast or IPTV can only be done by applying method c), which is based on exploring the real contents of the elementary streams carrying subtitles. When broadcasters provide coherent, synchronous and reliable contents in the SI and PSI tables, enhanced alternatives for measurement can be implemented.

A detailed procedure to measure the actual subtitling contents in DVB digital television or IPTV TV channels for subtitling time in television programs has been developed that can be found in (De Castro, 2010).

5 QUANTITY OF AUDIODESCRIPTION

During the research in Hermes-TDT project it has been found that, in the few television channels in Spain that broadcast audio description (channels owned by TVE or TV3), signalling is not in line with the DVB standard: the audio stream containing the audio description is signalled as if it were any of the other audio streams that convey alternate audio in other languages. This precludes any signalling based automated monitoring, and forces human intervention for any research on the quantity and quality of audiodescription in television channels.

6 CONCLUSIONS

As the provision of accessibility services is inseparable from the television content in future multimedia networks, legislative developments, driven by the need to support the definitive implementation of the accessibility to communication, should come together with the creation of supervision and control agencies that monitor the compliance with the obligations by involved parties. These entities need appropriate tools to carry out their function, an activity that, to evaluate service provision from both a quantitative and qualitative points of view, will also require appropriate analytical methodologies. A practical approach to do this has been presented in this paper. In particular, the procedure to measure actual subtitling times in TV broadcast is a requirement for a reliable monitoring of the accessibility services by TV providers, as quantity of subtitling are the fist *quality* parameters of the accessibility facilities offered. The results obtained can be used not only by the regulatory entities but also by the broadcasters as a measure of compliance and of the quality offered to the users. As the provision of accessibility services is inseparable from the television content in future multimedia networks, the described procedures can be used as a basis to monitor the level of subtitling in such services, as is being done in other areas (Fraga et al., 2009).

Discussion: technological developments may open a new door to the research in the field of accessibility, however, to find objective criteria to define quality indicators for it is a very complex task and not free of debate.

ACKNOWLEDGEMENTS

Hermes-TDT project has been funded by the Spanish I+D+i Avanza program and is participated by Universidad Carlos III de Madrid, Universidad Politécnica de Madrid, Universidad Rey Juan Carlos and Moviquity.

REFERENCES

Báez, I.C. & Fernández, S. (2010). “Spanish deaf people as recipients of closed captioning” In *Subtitle for the deaf and hard of hearing and audio description: First specific approaches and their application* (pp 25-44). Frankfurt: Peter Lang.

De Castro, M., Ruiz, B., Gonzalez, F. (2010), “Procedure to measure subtitling time in DTT and IPTV channels”; TEMU 2010, Int. Conf. on Telecommunications and Multimedia, Chania, Crete, Greece, 14-16 July 2010

European Telecommunications Standards Institute (ETSI) (2003), “Digital Video Broadcasting (DVB); Specification for conveying ITU-R System B Teletext in DVB bitstreams” (DVB-SI) ETSI EN 300 472 V1.3.1

European Telecommunications Standards Institute (ETSI) (2006), “Digital Video Broadcasting (DVB); Subtitling systems” (DVB-SUB 1.3.1) ETSI EN 300 743. V1.3.1

European Telecommunications Standards Institute (ETSI) (2009), “Digital Video Broadcasting (DVB); Specification for Service Information in DVB Systems,” (DVB-SI) ETSI EN 300 468 V1.11.1

Forbes, G. (2001), "Closed captioning transmission and display in digital television", *Proceedings of the Second International Workshop on Digital and Computational Video*, Vol., no., pp.126-131, 2001

Fraga, A, Paneda, XG, Garcia, R, Melendi, D. and Cabrero, S. (2009), “A Non-intrusive Estimation for High-Quality Internet TV Services”, *Proceedings of the Future Multimedia Networking*, in LNCS, Vol. 5630, pp.153-162, 2009

Huang, Y., Hsu, L. and Sandnes, F.-E. (2007), "An Intelligent Subtitle Detection Model for Locating Television Commercials," *IEEE Transactions on Systems, Man, and Cybernetics*, Part B: Cybernetics, Vol.37, no.2, pp.485-492

ISO/IEC (2007) "Information technology - Generic coding of moving pictures and associated audio information: Systems", ISO/IEC 13818-1:2007

Kayess, R. and French, P. (2008) “Out of Darkness into Light? Introducing the Convention on the Rights of

Persons with Disabilities", *Human Rights Law Review*, Nº 1, str. 1-34

Martín Edo, C.A., Jiménez, D., Cisneros, G. & Menéndez, J.M. (2007) "Implantación de la accesibilidad en la televisión digital: situación actual y futuros desarrollos". In Ruiz, B. & Utray, F. (eds.) *Accesibilidad a los medios audiovisuales para personas con discapacidad. AMADIS 06* (pp. 115-129). Madrid: Real Patronato sobre Discapacidad, 2007.

Mochizuki, Y., Oda, A., Miura, T., Mitsui, K., Kotaki, K., Kameda, K., Amano, Y., Nishida, Y. (2006), "Transport Stream Real-time Analyzer System", *Proceedings of the NAB Broadcast Engineering Conference*, pp.76-81

Ofcom (2008), "Code on Television Access Services", http://www.ofcom.org.uk/tv/ifi/codes/ctas/ctas.pdf (accessed 17 March 2010)

OSI/EU Monitoring and Advocacy Program (2005), "Television across Europe: regulation, policy and independence", *Open Society Institute*, Hungary, http://www.soros.org/initiatives/media/articles_publications/publications/eurotv_20051011 (accessed 20 March 2010)

Peng, C. and Vuorimaa,P. (2002), "Decoding of DVB Digital Television Subtitles", *Proceedings of the 20th International Multi-conference on Applied Informatics*, pp. 143-148

Pereira,A. (2010). Including Spanish Sign Language in subtitles for the deaf and hard of hearing. In *Subtitle for the deaf and hard of hearing and audio description: First specific approaches and their application* (pp. 103-115). Frankfurt: Peter Lang.

Pérez-Ugena, A. & Vizcaino, R. (2008). *Ulises y la comunidad sorda. Hacia el desarrollo de tecnologías comunicativas para la igualdad de oportunidades.* Madrid: Observatorio de las realidades sociales.

Reimers, U.H. (2006), "DVB-The Family of International Standards for Digital Video Broadcasting," *Proceedings of the IEEE*, vol.94, no.1, pp.173-182

Spanish Digital Television Technical Forum, WG 5 for accessibility (2006), "Accesibilidad en Televisión Digital para personas con discapacidad," ("Accessibility to the Digital Television for people with Disabilities"), Spanish Ministry of Industry, Tourism and Trade

Spanish Government (2010), "Ley General de la Comunicacion Audiovisual", last proposal published available at http://www.boe.es/boe/dias/2010/04/01/pdfs/BOE-A-2010-5292.pdf (accessed 30 June 2010)

Stienstra, A.J. (2006), "Technologies for DVB Services on the Internet", *Proceedings of the IEEE*, vol.94, no.1, pp.228-236

United Nations (2006), "Convention on the Rights of Persons with Disabilities", http://www.un.org/disabilities/convention/conventionfull.shtml (accessed 17 March 2010)

Yildirim, K.S., Ugur, A. And Kinaci, A.C. (2007), "Design and Implementation of a Software Presenting Information in DVB Subtitles in Various Forms", *IEEE Transactions on Consumer Electronics*, vol.53, no.4, pp.1656-1660

GNU RADIO
A New Paradigm for Software Defined Radio

Javier Rodas, José A. García-Naya, Carlos J. Escudero, Luis Castedo
Department of Electronics and Systems, University of A Coruña, Campus de Elviña, 8, 15071 A Coruña, Spain
{jrodas, jagarcia, escudero, luis}@udc.es

Keywords: Software Defined Radio, Real-Time Implementation, Wireless Communications.

Abstract: With the advent of GNU Radio, a revolution in the SDR technologies is foreseen that will greatly reduce development and deployment costs in terms of hardware as well as manpower. The availability of the Universal Software Radio Peripheral (USRP) and the USRP2, an extremely low-cost, mid-performance, real-time capable system equipped with a great variety of pluggable radiofrequency front-ends covering any frequency band from 50 MHz to 2.9 GHz and from 4.9 GHz to 5.9 GHz leads to a unique device all over the world. Furthermore, such hardware is complemented with the GNU Radio software, an open-source, flexible, sophisticated and powerful software toolkit specifically designed for SDR-based implementations and digital signal processing carried out in general-purpose processor. In conjunction, the GNU Radio software together with the URSP and, especially, the URSP2, constitute a new paradigm to be taken into account when SDR systems are to be developed.

1. INTRODUCTION

Software-Defined Radio (SDR) is defined by the SDR Forum as a "*radio in which some or all of the physical layer functions are software-defined*" (SDR Forum, 2010). Only radiofrequency (RF) and signal rate conversion operations are somehow implemented in hardware, while the remaining radio functionalities are implemented in software. SDR technologies are particularly suitable for implementing the multi-standard wireless communications equipments demanded by the Next Generation Mobile Networks (NGMN) alliance (NGMN, 2010).

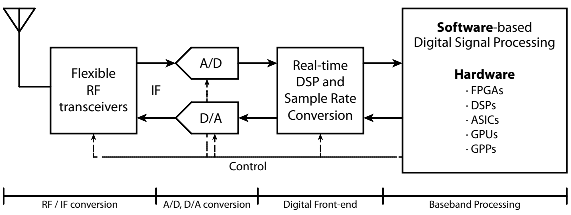


Figure 1: Block diagram of the main components of an SDR system.

Figure 1 shows a block diagram containing the basic components of an SDR system. On the right-hand side the baseband processing system is shown, which is in charge of carrying out all digital signal processing operations. Depending on the requirements of the SDR system, the hardware utilized can vary from ordinary general-purpose processor to real-time processors like Digital Signal Processors (DSPs) and Field Programmable Gate Arrays (FPGAs), and even application-specific integrated circuits (ASICs). In order to reduce the transfer rate between the DSP software and the Analog-to-Digital (A/D) and the Digital-to-Analog (D/A) conversion section, a digital front-end is included, which mainly carries out sample rate conversion operations (i.e., digital up and down conversions). Finally, flexible and user-configurable RF front-ends are utilized to be able to intercommunicate at different RF bands.

However, the commercial success of SDR solutions has been obstructed by the high cost and low flexibility of the currently available programmable devices (mainly those based on DSPs and/or FPGAs) (Hunt Engineering, 2010; Innovative Integration 2010; Lyrtech, 2010; Nallatech, 2010; National Instruments, 2010; Pentek, 2010; Signalion, 2010; Sundance, 2010). With the advent of GNU Radio (GNU Radio, 2010), a revolution in SDR technologies is foreseen that will greatly reduce development and deployment costs in terms of both hardware and manpower. This paradigm change is evidenced by the large number of GNU

Radio solutions that have been recently developed (CGRAN, 2010): GPS receiver, 802.11a/b/g/p and 802.15.4 applications, RDS FM receiver, RFID solutions, etc.

GNU Radio is a free and open software toolkit for the development of SDR technologies. GNU Radio provides the necessary blocks for implementing reconfigurable software radios using the Universal Software Radio Peripheral (USRP) hardware. The USRP (and its successor, the USRP2) is a low-cost external digital front-end and radiofrequency (RF) hardware manufactured by Ettus Research LLC (Ettus, 2010). It contains a programmable signal rate conversion before the digital front-end in order to reduce the required transfer rate with the host PC. RF front-ends are available as pluggable daughterboards. Consequently, the same motherboard can be used with a great variety of daughterboards, each one designed to operate in specific RF bands.

The GNU Radio and the USRP (including the USRP2) work with the following operating systems: any GNU/Linux distribution (with a 2.6 kernel version), on the Mac OS X (either in the Power PC or the x86 architectures), on the Microsoft systems (including Windows 2000, XP, Vista, and 7) and, finally, on FreeBSD and NetBSD UNIX systems.

Given that, first, the price of the USRP and the URSP2 is, respectively, 700 USD and 1400 USD; second, the price of the majority of the daughterbords range from 275 USD to 450 USD; and, third, the GNU Radio software is available for free, the total cost of GNU Radio solutions is very attractive, especially when compared to the cost of commercial solutions offered by companies like Lyrtech, Nallatech, National Instruments, or Sundance among others. The cost of such commercial solutions increases by a factor of ten or even more. Furthermore, commercial solutions are not based on free hardware designs and free software (GNU, 2010; GPL, 2010) and, thus, there are additional expenses for the software licenses.

In this paper, we provide a detailed overview of the GNU Radio hardware and software. More specifically, we focus our analysis on the URSP and the recently launched URSP2 hardware. We also stress the applicability of the URSP2 to the implementation of multiple-antenna, real-time, high-performance, broadband wireless communications systems at a low-cost. Finally, we also emphasize the opportunities provided by the GNU Radio hardware – including both USRP versions – as well as its main limitations.

2. GNU RADIO OVERVIEW

GNU Radio is a free software framework for developing SDR applications (GNU Radio, 2010), providing signal processing blocks as well as drivers to interact with the hardware. GNU Radio defines an application development model based on a set of blocks (the signal processing operations) connected on a flow graph basis. The blocks are implemented in highly optimized C++, specifically suitable for the x86 architecture that is, by far, the majority in the general-purpose processor industry. Such blocks are interconnected between them by means of Python, a high-level language suitable for purposes like this. With the above-mentioned approach, it turns out that it is very easy to translate a typical description of a wireless communications system into a design suitable for the GNU Radio framework.

Although the GNU Radio software perspective is very interesting, in this paper we focus on the hardware side. From now on, we elaborate on the Universal Software Radio Peripheral (USRP) and its successor, the USRP2.

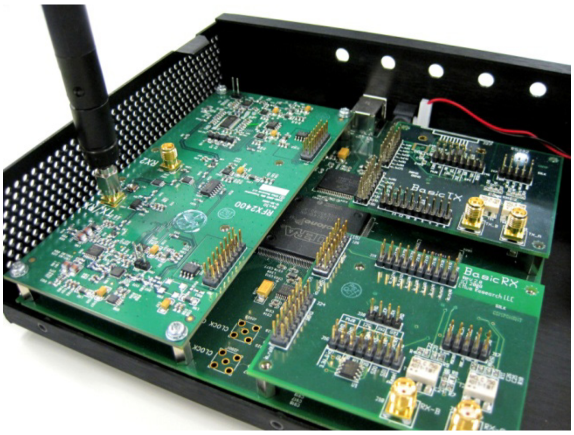


Figure 2: Picture of the USRP equipped with a RFX2400, a BasicRX and a BasicTX daughterboards.

3. THE USRP

Generally and simply speaking, the USRP (see Figure 2) is a low-cost piece of hardware that converts the analogue signals received by the antennas into discrete-time signals that are fed into an ordinary host to be processed by means of a general-purpose processor. Figure 3 shows a block diagram of the USRP hardware.

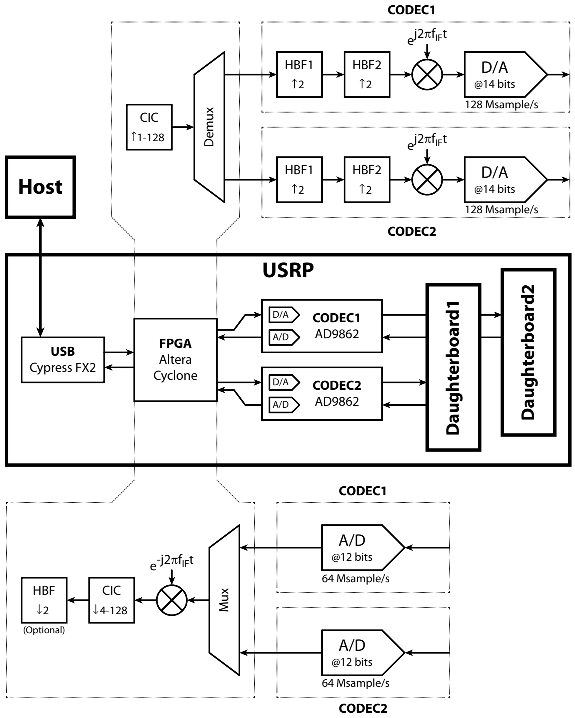


Figure 3: USRP block diagram. D/A, A/D converters as well as data path assume complex-valued signals.

Basically, the USRP consists of a motherboard plus up to four RF front-ends referred to as daughterboards (more on this in Section 5). The motherboard contains the following basic elements:

- A Cypress FX2 Universal Serial Bus (USB) controller (Keil, 2010) which handles the data transport between the host and the USRP motherboard.
- An Altera Cyclone EP1C12 FPGA (Altera, 2010) that carries out the data routing between the USB controller and the CODEC section. It also implements some digital signal processing operations for the user-selectable sample rate conversion and digital mixing to baseband at the receiver side.
- For the CODEC section, the USRP incorporates two Analog Devices AD9862 mixed-signal front-end processors (CODEC) directly attached to the FPGA (AD9862, 2010). The only transmit signal processing blocks included in the FPGA are the Cascaded Integrator-Comb (CIC) interpolators (Hogenauer, 1981). The USRP allows the interpolator output to be routed to any of the four CODEC inputs. The remaining steps to be performed at the transmitter (namely, the digital up conversion and the complex quadrature modulation) are carried out inside the AD9862. At the receiver side, however, only a half-band filter decimating by two is implemented inside the AD9862. The complex quadrature demodulation as well as the digital down conversion is carried out inside the FPGA.

Figure 4: Photograph of the USRP2 equipped with the RFX2400 daughterboard.

The default firmware included with the GNU Radio software configures the D/A converters to 128 Msample/s, while the A/D converters sample at 64 Msample/s. Given that the USB can sustain a rate equal to 32 MB/s and that complex baseband signals take four bytes for each complex-valued sample, it results in a maximum bandwidth of 8 MHz shared among all transmit and receive paths. Finally, Ettus declares a Spurious-Free Dynamic Range (SFDR) equal to 85 dB for the A/D converters and equal to 83 dB for the D/A converters.

4. THE USRP2

The USRP2 (see Figure 4) is the successor of the USRP, offering improved versions for the CODEC, the FPGA, and the data transfer system with the host, which is carried out through a Gigabit Ethernet connection (see Figure 5). Contrarily to the USRP, interpolation and decimation filters as well as digital mixing are all implemented in the Xilinx Spartan 3 FPGA, while the CODEC consists of an Analog Devices AD9777 dual D/A converter (AD9777, 2010) plus a Linear Technologies LTC2284 dual A/D converter (LTC2284, 2010), resulting in two D/A converters and two A/D converters, both configured to sample at 100 Msample/s, dealing with intermediate frequency (IF) signals. Additionally, the AD9777 implements a four-time interpolation filter in the analogue domain, simplifying the requirements for the analogue reconstruction filters and improving the signal quality at the TX side. The motherboard also incorporates a CPLD for loading the FPGA firmware from an ordinary Secure Digital (SD) card.

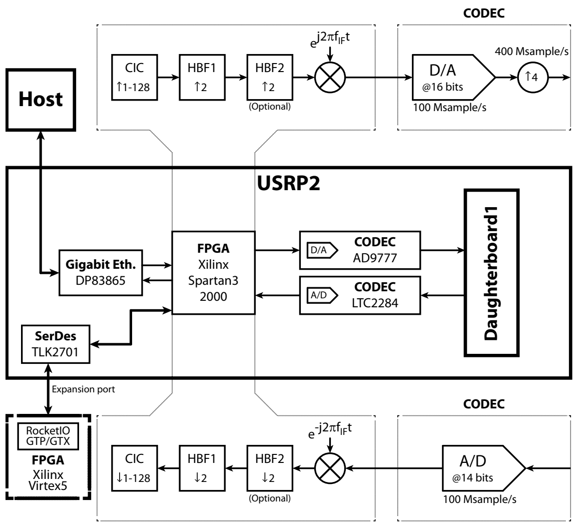


Figure 5: USRP2 block diagram. D/A, A/D converters as well as data path assume complex-valued signals.

The USRP2 is specified to be able to manage 50 MHz of bandwidth shared between the transmitter and the receiver sections with full-duplex operation, becoming the broadband wireless device with the lowest cost per bandwidth unit available in the market. However, it features a single CODEC, which leaded to the incorporation of mechanisms to coherently synchronize various USRP2 units by incorporating a user-configurable clock distribution system. Particularly, the USRP2 also includes a specific Multiple-Input Multiple-Output (MIMO) expansion port allowing coherent synchronization of two or four URSP2 units.

However, what makes the USRP2 a unique device from the SDR point of view is the possibility of attaching to it a high-performance FPGA (e.g. a Xilinx Virtex 5 by means of the Rocket I/O connection) utilizing a 2.2 Gbit/s high-speed Serializer/Deserializer (SerDes), which provides a tremendous increase in bandwidth with respect to previously available technologies. This opens the door for SDR-based, real-time, high-performance, multiple-antenna, broadband wireless communica-tions systems at a low-cost.

Finally, Ettus declares a SFDR of 88 dB for the A/D converters and at least 80 dB for the D/A converters.

5. RF DAUGHTERBOARDS

Ettus Research LLC provides three different sets of pluggable RF daughterboards. The first set is constituted by the so-called BasicTX, BasicRX, LFTX and LFRX, which permits direct access to the CODEC inputs and outputs (through coaxial cables), thus allowing connection with RF front-ends different than those provided by Ettus. The second set consists of a Very High Frequency (VHF)/Ultra High Frequency (UHF) receiver based on an analogue TV tuner, referred to as TVRX. Such set is completed with the DBSRX, an 800 MHz to 2.4 GHz receiver covering many frequency bands of interest such as, for example, Global Positioning System (GPS) and Galileo, Digital Enhanced Cordless Telecommunications (DECT) as well as some astronomy bands. Finally, the third set is formed by the so-called transceiver daughterboards, namely: WBX, RFX400, RFX900, RFX1200, RFX1800, RFX2200, RFX2400, and XCVR2450, covering the frequency bands from 50 MHz to 2.9 GHz and from 4.9 GHz to 5.9 GHz.

Table 1: RF transceiver daughterboards.

Model name	Frequency range (MHz)	TX power (dBm)	Price (USD)
WBX	50 - 2200	20	450
RFX900	750 - 1050	23	275
RFX1200	1150 - 1450	23	275
RFX1800	1500 - 2100	20	275
RFX2200	2000 - 2400	20	275
RFX2400	2300 - 2900	17	275
XCVR2450	2400 - 2500 4900 - 5900	20	400

Table 1 shows the main features exhibited by the transceiver daughterboards as well as their prices according to the official pricelist available at the Ettus webpage in June 2010. All transceiver daughterboards are MIMO capable and completely reconfigurable by software. They also incorporate a TX/RX switch, built-in Receive Signal Strength (RSS) measurement capability, 70 dB automatic gain controller (AGC), adjustable TX power and, finally, some of them can operate in a full-duplex way.

6. USRP VS USRP2

The USRP2 offers a considerable improvement with respect to the previous USRP. Hereafter, we stress the most significant differences:

Connection to the Host

The first difference is found in the interconnection mechanism with the host. In theory, the USRP2 can sustain up to 200 MB/s (50 mega complex sample/s) instead of the 32 MB/s sustained by the USB 2.0 in the USRP. However, different tests reported that only 100 MB/s can be sustained (without *underruns*), leading to a maximum bandwidth of 25 MHz. As an additional advantage, the Ethernet cable can be much longer than the equivalent USB 2.0 cable, thus distances longer than two meters from the host are possible.

CODEC

The USRP CODEC is based on the AD9862, while the URSP2 CODEC utilizes the AD9777 and the LTC2284, both offering more resolution bits and higher sampling rate. However, the sampling rate at the D/A converters is higher in the USRP1, but the AD9777 is capable of sampling up to 160 MHz and, together with its internal interpolation filter, generating a 320 Msample/s signal.

MIMO Capabilities

The USRP2 is ready to be coherently synchronized with other USRP2 units. The user can choose, by software, between an external reference and the internal one (generated utilizing a VXCO), which is also better than that included in the USRP. Both the USRP and the USRP2 support two TX/RX antennas in a MIMO system. However, the USRP2 is already prepared to construct a system with a greater number of antennas, keeping it coherently synchronized. According to Ettus specification, the USRP2 can form an eight-antenna coherent system (employing a four-way MIMO link device).

Real-time Capabilities

Last, but not least, the USRP2 includes the SerDes high-speed interconnection mechanism to be able to attach another high-performance FPGA system to the URSP2 FPGA, thus opening the door for the integration of the URP2 into high-performance, real-time SDR wireless communications systems. Note that here are no real-time capabilities with the USRP.

7. ADVANTAGES AND LIMITATIONS OF GNU RADIO

The main advantages offered by GNU Radio are summarized below:

- Low-cost solution. Compared to commercial solutions, GNU Radio is a free software platform and you only pay for the hardware manufacturing expenses.
- Low-cost software implementation. GNU Radio platform provides the engineer a very flexible and sophisticated SDR programming environment.
- Some parts of the hardware designs are open, thus easing customization to meet specific requirements.
- GNU Radio is platform-independent. The only requirement is a USB library (for the USRP) and it is available for the most used operating systems.
- The available set of pluggable daughterboards covers the RF bands ranging from 50 MHz to 2.9 GHz, and from 4.9 GHz to 5.9 GHz.
- Both URSP and USRP2 incorporate mechanisms to provide a common reference signal to be able to synchronize several units.

However, besides the above-mentioned advantages, GNU Radio also presents some drawbacks:

- Although Ettus claims that the USRP and the USRP2 are open design hardware, the sources of the hardware design are not available, neither from the Ettus, nor from the GNU Radio webpages. Instead, only the schematics are distributed, and they are generated with gEDA (gEDA, 2010), while the PCB layouts, designed with PADS (PADS, 2010), are not shared. Consequently, the software does follow the free source model, but the hardware design does not (at least not completely).
- GNU Radio is a sophisticated programming environment. However, a great amount of people specialist on digital signal processing is familiarized neither with C++ nor with Python, thus the knowledge curve has a greater slope than in the case of well-known tools like MATLAB (Mathworks, 2010).
- The first-generation USRP hardware employs a low-cost USB-based solution to solve the interconnection with the host. Consequently, the maximum sustainable sample rate is reduced to

8 Msample/s shared among all antennas. The USRP2 overcomes such limitation by replacing the USB connection with a Gigabit Ethernet.

- The USRP offers very limited possibilities for real-time implementations. However, the USRP2 incorporates a SerDes connection that solves this problem.

With respect to the SDR development, the GNU Radio together with the URSP2 constitutes a very economical solution in terms of costs and manpower. The availability of extremely low-cost hardware capable of handling signals with 25 MHz of bandwidth at any RF carrier from 50 MHz to 5.9 GHz constitutes by itself a unique solution all over the world. Furthermore, the integration of the interface and configuration of such hardware in the GNU Radio software, supported in the most used platforms, and available as open source constitutes the ideal complement for the hardware, making possible to transmit signals from an ordinary host and/or acquire them at low-cost and with very little time and effort.

The advent of the USRP2 and its connectivity with high-performance, real-time solutions such as, for example, the Xilinx Virtex 5, leads to a superb solution even for replacing current, expensive, real-time hardware products.

8. CONCLUSIONS

In this work we presented the GNU Radio software as well as the URSP and the USRP2 hardware platforms manufactured by Ettus Research LLC as a solution to be taken into account when SDR-based wireless communications systems are being devised. We introduced the main idea of GNU Radio and described in detail the USRP, the USRP2 and the broad set of pluggable RF daughterboards available. Next we compared the USRP with the USRP2, arriving at the conclusion that the USRP2 is a much more suitable device for high-performance SDR systems. Finally, we described the main advantages as well as the main limitations exhibited by GNU Radio when it is combined with the URSP and/or the USRP2.

ACKNOWLEDGEMENTS

This work has been partially supported by Xunta de Galicia, Ministry of Science and Innovation, Ministry of Industry, Tourism and Trade and FEDER funds from the European Union (07TIC019105PR, 09TIC008105PR, TEC2007-68020-C04-01, TSI-020301-2008-2, CSD2008-00010).

REFERENCES

AD9777, 2010, Analog Devices dual D/A Converter, URL: www.analog.com/en/digital-to-analog-converters/da-converters/ad9777/products/product.html

AD9862, 2010, Analog Devices Mixed Signal Front-End, URL: www.analog.com/en/rfif-components/rxtx-subsystems/ad9862/products/product.html

Altera, 2010, Altera Cyclone FPGA Family Overview, URL: www.altera.com/products/devices/cyclone/overview/cyc-overview.html

CGRAN, 2010, The Comprehensive GNU Radio Archive Network (CGRAN), URL: www.cgran.org/wiki/ProjectsGNU Radio, 2010, www.gnuradio.org

Ettus Research LLC, 2010, URL: www.ettus.com

gEDA, 2010, URL: www.gpleda.org/

GNU, 2010, The Free Software Definition by GNU, www.gnu.org/philosophy/free-sw.html

GPL, 2010, The GNU General Public License (GPL), www.gnu.org/licenses/gpl.html

Hogenauer, Eugene B., 1981, An Economical Class of Digital Filters for Decimation and Interpolation, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-29, no 2, April, 1981.

Hunt Engineering, 2010, URL: www.hunteng.co.uk

Innovative Integration, 2010, URL: www.innovative-dsp.com

Keil, 2010, Cypress FX2, URL: www.keil.com/dd/chip/3205.htm

LTC2284, 2010, Linear Technologies dual low-power ADC, URL: www.linear.com/pc/productDetail.jsp?navId=H0,C1,C1155,C1001,C1150,P15833

Lyrtech, 2010, URL: www.lyrtech.com

Mathworks, 2010, The Mathworks MATLAB, URL: www.mathworks.com/

Nallatech, 2010, URL: www.nallatech.com

National Instruments, 2010, URL: www.ni.com

NGMN, 2010, Next Generation Mobile Networks, URL: www.ngmn.org

PADS, 2010, Mentor Graphics PADS PCB Design Tools, URL: www.mentor.com/products/pcb-system-design/design-flows/pads/

Pentek, 2010, URL: www.pentek.com

SDR Forum, 2010, URL: www.sdrforum.org

Signalion, 2010, URL: www.signalion.com

Sundance, 2010, URL: www.sundance.com

The Free Software Definition by GNU, URL: www.gnu.org/philosophy/free-sw.html

The GNU General Public License (GPL), URL: www.gnu.org/licenses/gpl.html

SMART CITIES: GOING TOWARDS THE FUTURE

Josep Paradells Aspas*, Anna Calveras Augé*, Carles Gomez Montenegro*
Departament d'Enginyeria Telemàtica- Wireless Networks Group (WNG)
Escola Tècnica Superior de Telecomunicacions de Barcelona
Universitat Politècnica de Catalunya
C/ Jordi Girona, 1-3, 08034 Barcelona. Campus Nord UPC. Edifici B3. Spain
josep.paradells@entel.upc.edu, anna.calveras@entel.upc.edu, carlesgo@entel.upc.edu

Keywords: Smart Cities, Sensor Networks, Wireless Sensor Networks.

Abstract: The evolution of electronic technology has allowed creating small size devices with reduced process and communication capacity. This trend has been reflected in personal computers, mobile phones, and more recently in sensor networks. These sensor networks with all these elements are already being deployed. South Korea was a leading country in this type of solution, but others have followed suit. In Sant Vicenç dels Horts, a small town close to Barcelona for more than two years has experimented with a network of 80 nodes that provides traffic information, the temperature and humidity of the green areas, and location of buses. Other major cities have launched initiatives in the same direction as Santander and Barcelona to experience the real potential of this technology. Thanks to different collaborations with Fundació I2Cat, STSI and Cátedra RED.es-ETSETB-UPC several deployments and research is done in this area.

1 INTRODUCTION

The evolution of electronic technology has allowed creating small size devices with reduced process and communication capacity. This trend has been reflected in personal computers, mobile phones, and more recently in sensor networks. Such devices allow virtualization of the physical world and the interaction between the real world and its digital counterpart. A physical parameter such as soil moisture can be translated into digital information, processed and generally result in a command that can open a valve that allows water and in turn change the humidity level. This possibility is no longer an experiment and emerging in real applications. One of those who are raising more interest is the known as smart cities. The idea is to deploy around town sensor networks that capture relevant events and magnitudes for the proper management of the city. These sensors can capture information that can be stored, processed to identify situations requiring interaction or prepared for viewing by the citizen. Therefore, trying to capture relevant information and improve efficiency in the management and use of the city.

A system to carry out those functions is not simple. We cannot think that a sensor or a sensor network is sufficient. One must think in the capture of different types of information from the humidity, temperature, occupancy of a container, the presence of a pedestrian or car at some point, etc. These transducers must deliver information to a processor that must formatting and transmitting information via radio to a central element. In order to minimize battery consumption and the costs of network deployment, the sensor nodes themselves do as node routers. The information of a node jumps from node to node until it reaches its destination. This technique known as multi-hop also offers additional auto configuration capacity and improves network availability as in case of failure of a node it can be found an alternative way using other nodes.

The information from a sensor node can be directed to a sensor node to a sink node or gateway that connects the sensor network with the traditional fixed network. Once there, the information travels to a node where it is processed and stored. To facilitate the use of a sensor network by various third parties is possible to define open interfaces and use of data formats that are semantically rich.

These sensor networks with all these elements are already being deployed. South Korea was a leading country in this type of solution, but others have followed suit. In Sant Vicenç dels Horts, a small town close to Barcelona for more than two years has experimented with a network of 60 nodes that provides traffic information, the dampness of the green areas, temperature and humidity and location of buses. Other major cities have launched initiatives in the same direction as Santander and Barcelona to experience the real potential of this technology.

The paper shows the technological elements involved in creating an urban wireless sensor networks. It starts presenting the so called wireless sensor nodes and the type of transducer employed according the application. A section is devoted to present the whole architecture of the system describing the protocols involved in the communication. The paper ends with the description of the experimental networks installed in Sant Vicenç dels Horts and some conclusions earned.

2 SENSOR NODE PLATFORMS

Sensor networks have been studied in the last decade, but in recent years are emerging as a reality in some certain fields of use. One is the sensing of cities in what is known as "Smart Cities". This idea was promoted in South Korea under the name U-Cities and has carried out various experiments in different cities using sensor networks and RFID (Radio Frequency Identification) components. Cities are an excellent field of application of sensor networks. There are many elements to monitor and manage in an intelligent manner of a city should allow a reduction of costs and/or improving services to citizens.

Sensor networks [1] are based on nodes which have sensors that can sense a physical parameter and/or actuators, which allow you to interact with the physical environment. These nodes have a processor that can process the measure and transmit via radio to any other node or a sink where the data is injected into the network and sent to the destination. The processing capabilities of the nodes have evolved rapidly. From 8-bit processors (such as the Chipcon 2430) four years ago to 32bit they have today (such as Freescale MC13224V or Jennic JN5148) with a consumption not increased. As for radio interfaces proprietary solutions and standardized can be found using the ISM (The industrial, scientific and medical) bands at 868MHz and 2.4GHz. In principle the proprietary architectures offer better performance for a specific application, but thanks to economies of scale standardized solutions based on IEEE802.15.4 are emerging on the market of integrated circuits with outstanding characteristics. They offer transmitted power up to 100mW with a 150mA consumption and bits rates of 250kbps.

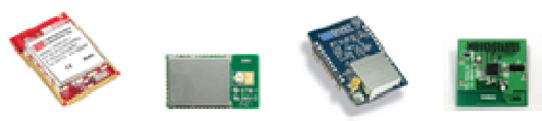


Figure 1: Wireless sensor node platforms. From left to right: FreeStar Pro from CEL based on Freescale MC13224V, JN5148 Module from Jennic, M2510 from Dust Networks and module from Huanor based on CC2430.

Combining radio communications and the availability of a processor in the node it is possible to build a multi-hop communication mechanism. Nodes are not limited to form a star topology network with one sink at the hub. They can build a multihop network where nodes collaborate among them to move the data from the source to any available sink. This alternative is more efficient in terms of power consumption and in providing reliability. With this type of network, in case of node or link failure it is possible to find an alternative path to reach de destination. In fact the network should be self-configuring, so that it should be able to be created and repaired without human intervention. This is an advantage in terms of deployment and maintenance since reduces significantly operating. A key issue underlying this concept is self-powering. Extending the idea of being self configuration means node running with batteries of with energy harvesting mechanisms that collect energy form the surrounding and store it in a battery. The cities are very attractive in terms or powering since it is quite common to have an electrical power supply or to benefit from other sources such light, noise, vibration, radiofrequency or pressure. When powered with battery it is required a large battery life, close to the technological obsolescence of the equipment. The common considered battery lives are between 5 and 10 years.

There is now a wide range of components that allows building a network of sensors. Manufacturers such as Texas Instruments, Atmel, Freescale, Jennic or Nordic Semiconductors (to mention some of them) offer MCUs and radio transceivers. With these elements a larger number of manufacturers have developed a set of modules. Among the most prominent we can mention MeshNetics, Crossbow, Radiocrafts, ArchRoc, Digi, Dust Networks or spanish Zolertia and Libelium. These platforms offer different benefits and are controlled by different operating systems. Some use free and open solutions such as TinyOS, Contiki or FreeRTOS while other solutions that the manufacturer has developed. It has to be noticed the variety of platforms and operating systems that we have at our disposal. This is an advantage, but both a complication to the case that we want to build a network in which certain that combine different elements. In this case it is necessary to have a functional layer in the form of middlleware that can help to mitigate the difficulties associated with the heterogeneity of network nodes..

Wireless sensor nodes incorporate one or several transducers. They can be used as a sensor to convert a physical magnitude to its digital representation or as an actuator, to convert a digital signal to a physical one. Actuators may be a lighting switch for a lamps or an electro-valve of an irrigation system. The variety of sensors is quite large and can be used for data logging or to control actuators. For example depending on various parameters such as level of soil moisture, temperature, solar radiation level or the presence of people you can activate an irrigation system in landscaped gardens. The result is an efficient use of irrigation water, maximizing the absorption of water on earth and avoiding any damage as a result of over watering. The following section describes variety of possible sensors and application that may result.

3 SMART CITY APPLICATIONS

The used range of sensors is very wide and is only limited by the applications which finally decide to install. As a review we will use a description based on applications and detailing the type of used sensor.

- **Lighting Control**. Use light sensors to monitor the brightness level of the street and turn on or off street lighting [2, 3].
- **Climatic conditions**. The temperature, humidity, atmospheric pressure, direction and force of wind and rainfall is an interesting point when planning the clothing, activity to do or how long it can take the journey to the office. This information has no significant variation in close areas and therefore the number of sensors can be very small in a city.
- **Green zones irrigation control**. The green areas are usually scarce in the city and therefore should be carefully maintained. The sensors are used more to decide the necessity of avoiding watering when the soil is already moist which is an unnecessary wastage of water and a possible plant damage. Temperature and even people presence can complement the moisture level sensor when deciding the moment of watering the plants.
- **Monitoring of air pollution**. The control of air quality requires several sensors for measuring the concentration of certain gases such as CO_2, SO_2 or NOx, presence of suspended particles or the concentration levels of pollen. In urban areas close to places with industrial activity or logistical exchangers may be interesting to have a broader range of sensors to detect the presence of leaks that could be harmful to the population.
- **Monitoring of noise pollution**. Road traffic, industry or leisure activities can produce noise levels that do not combine with activities that require some concentration such working or sleeping. Any noise excess should be identified. For this application it is possible to use microphones as sensors to measure noise levels.
- **Public Safety**. Public spaces can be damaged or improperly occupied. Sensors that can detect presence of people such PIR (Passive InfraRed), light level or noise level may indicate that some space is occupied by humans, and thus detect a security problem. Motion sensors also can detect if an object has been removed from its location.
- **Buried underground infrastructure**. The city has a lot of their infrastructure such as pipelines, telecommunications, water, effluent, gas or electricity under the ground. This site makes it difficult to monitor and when a problem becomes evident is too late. To carry out preventive monitoring sensors can be installed to detect gas or water spills or even fire. Another problem is the usage of the sewer network to eliminate toxic residues. The possibility to analyze the composition of sewer waters close to its source will allow localizing the source of these illegal dumps.

- **Traffic**. Timely and updated information on traffic congestion is very interesting to plan travel time and route. The flow of vehicles can be monitored easily with magnetic field sensors and vehicle density can be measured with ultrasonic sensors [4, 5, 6].
- **Parking.** The search for a parking space once we reached our destination causes a waste of time and additional fuel consumption. Updated information on the presence of vacancies can help reduce this expense. In order to monitor these vacancies we can use various techniques such as magnetic field sensors buried in the ground or ultrasonic sensors [7, 8, 9].
- **Rubbish Recollection**. The collection of waste, in particular the selective, forces a set of specialized vehicles by type of waste regularly traverse the city emptied containers which in some cases are almost empty and in other cases overflowed. Provide information on the filling level would allow for more efficient and rapid collection. The filling level of the containers can be performed with ultrasound systems that measure distance or amount of empty space.
- **Presence**. The occupation of public roads by pedestrians is highly variable. In some cases it is interesting to know whether an area is occupied by someone. For example you can control lighting levels depending on the presence of people in a marquee public transport or in the street itself. Advertising can be sensitive to the presence of people or control of the control lights (traffic lights) can change regime depending on the amount of people waiting to cross the street. You can use PIR sensors, pressure or electric field to determine the presence of people.
- **Management of public transport**. The moment of passage from the bus for one stop or occupancy of a bus allows passengers to plan their routes dynamically. In general, the information presented in the current system is not of an adequate precision in time and present information about the occupation. The availability of a network of sensors on the road to allow public to locate the bus to within tens of meters and counting systems that are in the bus can provide information about the level of occupation.

Figure 2: Example of a sensor board from Waspmote to measure concentration of Gases (CO, CO_2, CH_4, H_2S, NH_3, NO_2, ..[10]

Figure 3: Example of a sensor board from Waspmote to measure luminosity, tilt, weight, vibration, liquid level, hall effect, presence (PIR),..[11]

4 WIRELESS SENSOR NETWORK

The wireless sensor network is made of sensor nodes and gateways. Even the sensor nodes can be classified between nodes able to route packets and nodes able only to receive and transmit packets. In some literature this distinction appears as Full Function Device (FFD) and Reduced Function Device (RFD) respectively. The main distinction is the set of protocols they are using. Mainly in a sensor network there are protocol layers. The lower one consists on the physical and the medium access control (MAC). The MAC supports the sharing of the wireless media among different nodes. The IEEE802.15.4 is the most widely accepted standard for these commented functionalities. It can be used with several frequency bands according the local regulation. For example in Europe it is possible to use one single channels supporting 20kbps at 868MHz and 16 channels supporting 250kbps at 2,4GHz. As the last frequency band correspond to the same used by Bluetooth and Wi-Fi this band can be quite congested in terms of interferences. It is very useful to have several channels.

To support the movement of data between FFD it is needed a routing protocol. There a large number of alternatives, but the one used nowadays is based on an ad-hoc routing protocol called AODV. There is another one, called RPL that will become the clear alternative in situations with large number of nodes reporting and received data from a sink [12]. Data can be routed as MAC frames or as Internet packets. The first option results the most efficient but the working group 6LoWPAN from the IETF is working towards a nice integration between Internet and wireless sensor networks. As a result of this effort is gaining support the second option to facilitate the interworking between networks. According to this working group it is needed an adaptation layer between IEEE802.15.4 MAC frames and IP packets recognized as the 6LoWPAN layer. On top of the network layer is required a transport layer to provide end to end reliability and congestion control. The ZigBee proposal [13] is a complete set of protocols that reuses some existing ones, such the IEEE802.15.4, and defines new ones. For example they propose to use a routing protocol based on AODV and define a transport layer named Application Support Sublayer (APS). On top of that it is possible to locate the application layer directly as ZigBee or use a session protocol following the working group of the IETF, CORE (Constrained RESTful Environments). ZigBee uses the concept of application profile as a mechanism to define applications for specific uses. There are a set of defined applications profiles such the one for metering, health or building automation. In the last months it has been a trend in favour of a convergence between the IETF proposals and the ones from ZigBee. The idea is to reuse protocols from IETF with applications defined by ZigBee.

The information generated by the sensor is coded in an efficient way in order to minimise the required bandwidth, but once the data is delivered at a central platform to be stored and processed the information should be enhanced semantically providing information about data representation, location and a description. There are suitable formats such EEML[14] or SensorML [15].

In addition the processing of the measured data, the central platform may help in programming sensor nodes and in managing them. These general functions of support are associated to a middleware functionality that hides the complexity of the wireless sensor network to the user.

Information collected is stored in a data base for offline and online analysis. The first one has the objective to detect trends in the city evolution and solve a past problem thanks to the historical data. The online processing is used to detect situations demanding short term response. In the case of city management this means within a delay of seconds for a security problem or days for repairing a lamp pots for example. It is a general claim among the citizens to make the collected data available to the community. This data is useful to show the municipality is aware of what is happening in the city and also opens the possibility to exploit these data by third parties to build services. Also the data collected can be organised to be presented through a web interface in an automatic way so web contends are updated without human intervention.

5 EXPERIENCES

The ideas described in this paper were elaborated some years in the past and they have been implemented experimentally in the small town of Sant Vicenç dels Horts, near Barcelona. The experimental network consists of a network of around 80 nodes covering a bus line, its surroundings and some isolated spots. The system has a total of 3 gateways, two using an Ethernet interface to the Internet and another one with a cellular connection using GPRS/3G. Each gateway provides Internet connectivity to a part of the wireless sensors, but even with this sharing the network has multihop paths longer than 19 nodes. The network has FFD nodes that are active continuously and a number of RFD that only become active when they have to send data. FFD should be powered almost continuously. The adopted solution uses the power of the lamp post during the night to recharge a battery that keeps powering the sensor the rest of the day.

Figure 4: Detail of a sensor node located on a lamp post

This approach facilitates the network deployment and allows building a backhaul network able to provide uninterrupted connectivity. Lamp post nodes has two sensor (temperature and relative humidity), but most of the sensor are implemented in RFD such moisture sensor to decide the irrigation instant, sensor nodes located inside the rubbish container to indicate the degree of fullness, nodes using ultrasounds to measure if a parking slot is full or empty and traffic estimators based on the occupancy of the street.

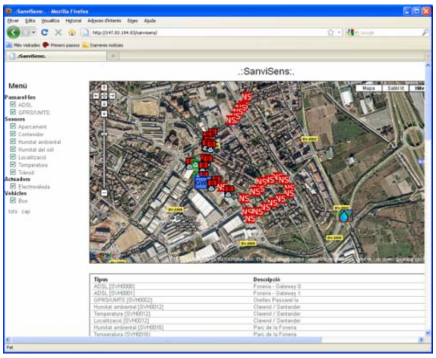


Figure 5: Web interface with information about the type and location of the different sensor nodes

Also bus has sensor node that informs about the sensor node from the backhaul closest to it. Knowing the location of each backhaul node it is possible to estimate the location of the bus at any moment. All the information obtained by the sensor is stored in a database and some of the data published using a Web server. This information is useful for the municipality and for the citizen.

6 CONCLUSIONS

Wireless sensor networks are a technology that has been evolving during the last decade and now it the time to start applying it in real deployments. As the technology is quite new at this point only specific applications appears as feasible usages. The building of smart cities seems one of the most promising. Reducing the management costs and providing more information to the citizen are among the priorities of the municipality managers. The technological platforms are already in the market. Certain degree of standardization is needed to eliminate uncertainty and assure fully Internet connectivity. In different places around the world are appearing wireless sensor networks for specific applications (parking slots management, pollution control…) but a general purpose network will be needed to get the most of this technology. The experience on Sant Vicenç dels Horts is an early example of this idea.

ACKNOWLEDGEMENTS

Thanks to the support of different organizations such Fundació I2Cat, the Municipality of Sant Vicenç dels Horts, the Secretaria de Telecomunicacions i la Societat de la Informació de la Generalitat de Catalunya, the Cátedra RED.es-ETSETB-UPC and the research project TEC2009-11453.

REFERENCES

[1] C. Gomez, J. Paradells, J. E. Caballero, “Sensors Everywhere: Wireless Network Technologies and Solutions”, Fundación Vodafone Spain. 2010.

[2] Chunguo Jing; Dongmei Shu; Deying Gu, “Design of Streetlight Monitoring and Control System Based on Wireless Sensor Networks”, Industrial Electronics and Applications, 2007. ICIEA 2007. 2nd IEEE Conference on, 23-25 May 2007 Page(s):57 – 62

[3] http://www.simplicityevent.philips.com/global/tomorrow/light_blossom/

[4] Special Session: “Intelligent Transportation and Traffic Telematics - Challenges, Applications, Standardization” 19th Annual IEEE International Symposium on PIMRC 2008.

[5] M. Nekovee, “Sensor networks on the road: the promises and challenges of vehicular ad hoc networks and grids”.

[6] T. Gruber, E. Schoitsch, “Automotive Visions beyond In-Car Driver Assistance: Integrated Traffic Management with Coopers”. http://ercim-news.ercim.eu/coopers-automotive-visions-beyond-in-car-driver-assistance

[7] Smart Dust project: http://robotics.eecs.berkeley.edu/~pister/SmartDust/

[8] Jung-Wook Lee; Yoon-Bong Yoo; Jae-Jeung Rho; Sae-Sol Choi, “An enhanced parking lot service model using wireless sensor network”, Industrial Informatics, 2008. INDIN 2008. 6th IEEE International Conference on, 13-16 July 2008 Page(s):349 – 354

[9] http://www.streetlinenetworks.com/site/index.php

[10] Gases Board: Technical Guide. Waspmote. Libelium. http://www.libelium.com/documentation/waspmote/gases-sensor-board_eng.pdf

[11] Events Board: Technical Guide. Waspmote. Libelium. http://www.libelium.com/documentation/waspmote/events-sensor-board_eng.pdf

[12] RPL http://tools.ietf.org/html/draft-ietf-roll-rpl-10

[13] ZigBee Alliance. http://www.zigbee.org

[14] Extended Environments Markup Language (EEML) Web site. http://www.eeml.org/

[15] SernsorML at OCG Web site: http://www.ogcnetwork.net/SensorML

REALISTIC COST ESTIMATION OF AN INTELLIGENT TRANSPORTATION SYSTEM ROLL-OUT

Erik Vanhauwaert, Sofie Verbrugge, Wim Vandenberghe, Mario Pickavet, Piet Demeester
Dept. of Information Technology (INTEC), Ghent University/IBBT, Gaston Crommenberglaan 8, 9050 Gent, Belgium
{erik.vanhauwaert, sofie.verbrugge, wim.vandenberghe, mario.pickavet, piet.demeester}@intec.ugent.be

Keywords: Intelligent Transportation System (ITS); roll-out; cost modelling; safety and recreational applications.

Abstract: This paper investigates the costs of rolling out an Intelligent Transportation System (ITS). Our cost model uses a set of possible technologies and applications, and can be further tailored to the most relevant scenarios. Unlike other research, this model investigates a joint roll-out, instead of separate applications. The reuse of existing infrastructure is also taken into account. This results in a more realistic cost estimation. Our model provides a cost break-up to identify the most crucial parts of the system in terms of costs. The case of the Belgian highways is used as a practical example.

1. INTRODUCTION

In essence, the concept of Intelligent Transportation Systems (ITS) implies the addition of information and communication technology to transport infrastructure and vehicles. A wide range of applications can run on this platform. For example, emergency services could automatically and immediately be informed when an accident happens (eCall application). Obstacle and Collision Warning can warn drivers for imminent collisions. More examples are discussed in the next chapter. In general, the intended benefits are increased traffic safety and efficiency and hence a positive impact on the environment through less traffic congestion.

The roll-out of such a system may require substantial investments in infrastructure. Moreover, many parties are involved, such as car manufacturers, network operators and traffic regulators. They need to be able and willing to cooperate and fairly allocate all costs and revenues. Here, we try and get detailed information on the required costs. There is little economic research on ITS costs so far, and it often focuses on a specific application (e.g. [1], [2]). A lot of this work is included on the Research and Innovative Technology Administration ITS overview site [3], but it provides building blocks rather than a complete picture. Some costs can be shared by many applications, which lowers the cost per application. Offering many services could also increase the users' willingness to pay (an important issue, according to [4]). Therefore, it is useful to consider a joint roll-out for a set of applications. To our knowledge, eIMPACT is the only (publically available) research project in which a similar investigation is made [5], though little information is given on what costs they have calculated and how. Thorough and complete cost information on the roll-out of an ITS appears scarce.

In this paper, we focus on cooperative applications, which require communication with infrastructure, unlike autonomous applications, which can be developed by manufacturers independently and are already on the market in different forms (e.g. ABS, ESP). We limit the roll-out to highways, thus omitting urban environments. This way, we group a set of applications that rely on a common infrastructure (the network) and we can investigate the impact of sharing general costs amongst all applications. To make our cost estimation more realistic, we will take re-use of existing network infrastructure into account. As a practical case, we've used the information of the Belgian highways and mobile networks.

In the next chapter, we define our roll-out scenario by selecting technologies and applications. Chapter 3 elaborates on our cost-model. In chapter 4, the results are discussed and we present our conclusions in the final chapter.

2. ROLL-OUT SCENARIOS

It is still unclear what the best roll-out scenario would be, and moreover, this is dependent on many factors, such as requirements of the end-users and the government, as well as technological and economical possibilities. Hence, the cost model is built to be flexible and allows many scenarios by configuring several building blocks.

2.1 Applications

The goal is to launch applications and they indicate our other requirements. However, there is no consensus on which applications should be rolled out and which shouldn't. This can partially be explained by the multitude of actors involved (e.g. driver preferences are different from technological feasibilities and government priorities). We've made a broad selection, based mostly on the work by [3], [6] and [7]. On top of safety applications, we've added recreational applications, because this could increase users' willingness to pay for the system. Our set of safety applications consists of eCall, Emergency Vehicle, Frontal Collision Avoidance, GPS Map Updates, Remote Diagnostics, Road Charging, Road Condition, Traffic Information and Traffic Management. The recreational application set contains internet browsing, audio and video streaming.

We describe applications by setting the most relevant parameters. The first is the percentage of drivers, who have a vehicle equipped with ITS technology, that will actually use the application. This is set to 100% for safety applications; for recreational applications, we estimate this will be 5% to 10%. The second is bandwidth use, and was estimated by analysing the size and frequency of messages. Finally, the complexity of the application translates to a development and maintenance cost (mainly personnel cost), as well as a server load.

In the model, we distinguish four options:

- only safety applications,
- only recreational applications,
- both,
- both, where the usage of recreational applications is set to 20%.

The last option is an extra test to check how future-proof the network is (as bandwidth demand continuously increases).

2.2 Network Technologies

In order to support our application set, vehicles will contain an On-Board Unit (OBU), which can communicate with Road-Side Units (RSU) or with the back-end/internet through an access point. The back-end is a control centre that runs and monitors applications centrally (Figure 1). This high-level architecture is in line with the European ITS Communication Architecture [8]. The network must meet the connection requirements of the applications. In this case, we distinguish three types of wireless access technologies:

- wireless medium to long-range unicast,
- wireless medium to long-range broadcast,
- wireless short-range unicast and geocast.

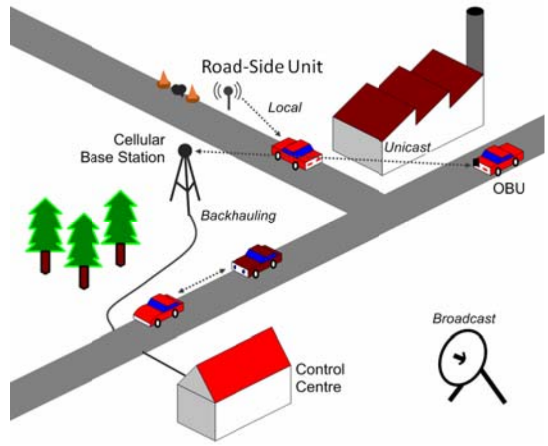


Figure 1: ITS components overview.

In case of long-range broadcast, contrary to unicast, bidirectional communication is not possible. It could be useful for preserving bandwidth in the network if all users require the same data (e.g. traffic information). Short-range communication can be cheaper than longer range, because it typically operates in unlicensed spectrum. But it is not always feasible (it would require too many RSU's). Fixed access, such as fiber, is only relevant as a backhauling solution, since OBU's obviously can't have a fixed connection.

For short-range communication, we only use CALM-M5 (this is the European version of IEEE 802.11p [9], also known as ITS-G5). CEN Dedicated Short Range Communication (DSRC) has a range of only 15m and does not support vehicle-to-vehicle communication (due to passive vehicle transporters). CALM-IR is based on directional infra-red communication and would thus require an expensive gantry over the road with one transceiver for each lane [8].

For broadcast we consider two options: MBMS or DVB-T. MBMS is a broad- and multicasting technique that can be used on top of UMTS and HSPA hardware. With Terrestrial Digital Video Broadcasting (DVB-T), traffic information could be sent to all vehicles using the DVB specifications for IP datacasting [10]. DAB was omitted, because it is far more limited in bandwidth, while the transceivers appear considerably more expensive. The use of a broadcast technology is optional (the medium-to-long range unicast network can also send messages to each vehicle).

Finally, for medium- to long-range unicast, we consider HSPA, mobile WiMAX and LTE. Older technologies would only support the most basic of applications, while HSPA is currently rolled out by almost all network operators that provide UMTS [11]. We assume HSPA will be the standard by the time an actual ITS roll-out takes place.

For the broadcast and local communication technologies, we used the standard specifications performance indicators, because it is unlikely that the ITS' requirements will exceed its capabilities. For the medium to long-range unicast technologies, we must take into account that the standard specifications are often not realizable in practice. The Erceg C path loss model shows that the realistic communication range for unicast technologies is a lot smaller than the standard indicators. For example, for mobile WiMAX, we take a range of 450m into account (inter-site distance 900m).

2.3 Other Building Blocks

Other relevant factors in determining the scenario are the adoption rate, the roll-out speed and the required availability of the services.

The adoption of the system is difficult to estimate, as there has never been an actual roll-out to compare with. Also, it is unclear whether the system should be enforced (e.g. obligatory integration in all new cars) or not. We've considered three possible adoption rates (Figure 2). Enforced adoption is based on the sales of new cars. The other options are modelled by a typical Compertz-curve; one version assuming a high acceptance, the other a low one.

We should also determine the speed of the roll-out. When an ITS is launched, it will not be used by everyone at all times right from the start. There can be a (long) transition period. There is an opportunity to save costs by postponing parts of the installation until the initial infrastructure is no longer sufficient. In Figure 2, a three-phased roll-out is assumed and indicated by the vertical bars. The initial installation is sufficient to cover an adoption rate of 30%, in case of enforced adoption. After three years, an additional installation is done to cover 60%. Finally, in year 8, the infrastructure is expanded to cover an adoption rate of 100%. Up to five phases can be defined in the model.

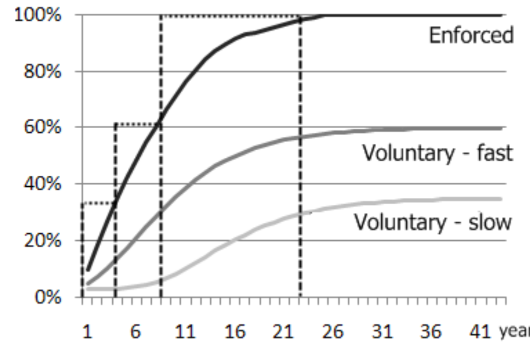


Figure 2: adoption scenarios, with indication of a three-phased roll-out for the enforced option.

The amount of bandwidth required in a base station will depend on the number of connected vehicles to it, and thus on the traffic situation. The worst case is one where there is a traffic jam in both directions. The odds of this happening, depends greatly on the exact location and it may not be necessary that the system can handle such extreme traffic circumstances everywhere. We distinguish five traffic circumstances with an increasing vehicle density and thus increasing communication requirements: (i) no traffic, (ii) sparse traffic, (iii) regular traffic, (iv) a one-way traffic jam and (v) a two-way traffic jam. The model allows defining the percentage of the highway that is equipped for each circumstance. This can be different for each phase, i.e. to allow an incremental increase.

3. COST MODEL

3.1 Capacity Demand

For any scenario, we need to know how much bandwidth and how many connections are required. Based on the application set and its parameters, we can calculate the average bandwidth consumed by one vehicle. In order to calculate the network load, we also need to know the vehicle density. We calculate the density for each of the five traffic situations. In case of no or sparse traffic, the network load is irrelevant, because the equipment is independent of the bandwidth usage (none in case of no traffic, just enough to have full coverage in case of sparse traffic). For regular traffic, we calculate the number of vehicles on one lane by making a few simplifying assumptions. We consider all vehicles to drive exactly the maximum speed, with two second intervals (based on the rule of thumb for safe driving). In this case a simple formula calculates the number of vehicles with average length *a* and with an average of *v* meters per second on one lane of a piece of highway with length *L*:

$$Number\ of\ vehicles = \frac{L}{a + 2.\mathrm{v}}$$

This is then multiplied by the number of lanes. In the model, this is slightly enhanced by also taking into account the percentage of trucks, their lower maximum speed and higher average length. While this model obviously simplifies reality, the calculated 150 vehicles/km on a 6-lane highway are in line with previous work that determined vehicle density by traffic measurements provided by the Flemish government [12].

In case of traffic congestion, a different model is used. Here we assume that vehicles move in waves: one vehicle fills the empty spot in front of itself, followed by the next vehicle. This way, only a percentage *p* of vehicles is moving. Non-moving vehicles have a small distance d_{still} between each other, while moving vehicles are driving when there is a larger distance d_{drive} in front of them. Again, a simple formula follows (which is again corrected for the percentage of trucks on the highway):

$$Number\ of\ vehicles = \frac{L}{a + d_{still} + p.d_{drive}}$$

This leads to an estimate of 631 vehicles/km on a 6-way highway in case of a two-way traffic jam. In case of a one-way traffic jam, we combine three lanes of regular traffic and three lanes of congested traffic, or 390 vehicles/km.

The network load on the highway is then simply the number of vehicles on that part, times the average bandwidth required.

3.2 Network Dimensioning

The network needs to be designed such that it can meet the capacity demand at the lowest cost. The input information is: (a) the capacity demand, (b) the distribution of existing sites and to what extent they can be reused, and (c) the technological parameters. Dividing the available bandwidth by the capacity demand gives us the required site density. However, the highway is not uniform and neither is the existing network site distribution. Information on traffic jams and base station locations is publicly available in Belgium. A visual check reveals that the densest regions of existing sites mostly overlap with the busiest parts of the highway. We assume that the overlap is complete, as this greatly simplifies the calculation. The total number of sites available next to the highway can be obtained by combining GIS-data of the highways and the network sites. There are about 650 sites located within 200m of the highway. Conceptually, the model calculates the number of new sites as follows. First, we split the highway into many small parts (the model works continuously, not discrete), and sort these from least to most busy. The busiest parts of the highway will also require equipment to handle the heaviest traffic situations. The capacity demand we calculated before can thus also be sorted from low to high and mapped on each part of the highway. Similarly, we also map the existing site distribution on it. Basically, we now have the number of required sites and the number of existing sites, for each part of the highway. Subtracting one from the other gives us the number of new base stations and the required equipment to connect them.

3.3 Cost Overview

In order to get a view on all costs, we consider the different phases in the life of the project and the different aspects of each (Figure 3).

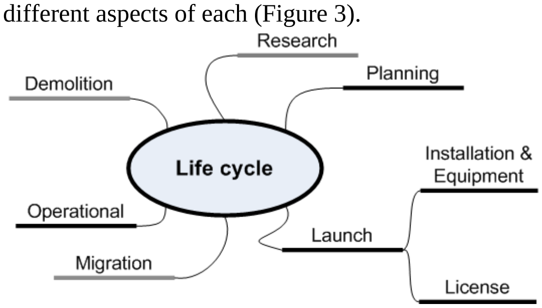


Figure 3: Overview of phases in ITS life cycle.

The planning and research phase can be a bit hard to differentiate from each other. Moreover, research costs by private companies will be integrated in the price of the final products and services they offer. Therefore, we only take into account the cost of planning the locations of new sites and transceivers. This is calculated by using the number of network planners as a cost driver.

The installation and material costs are part of the launch of the system, but they are spread in time due to the phased roll-out. The new network infrastructure leads to a bill of material, and installation and transportation costs can be estimated by using distance driven and hours spent as cost drivers. The existing network, such as base stations and backhauling connections, can be re-used, but we assume that only 40% of the transceivers will actually be available, because that network is also used to serve cell phone users. Lease costs are taken into account, estimated at about 10% of the acquisition cost per year.

The migration (or connection) cost is simply the installation of an OBU in this case. We assume the unit itself will cost about 200EUR (based on building blocks and also similar to Personal Navigation Devices). Aftermarket installation could prove difficult: a connection must be made to the internal CAN-bus and an antenna must be placed outside the vehicle body for good reception. There is little information available on this point; we assume a fixed fee of 100EUR for installation.

The operational phase contains many elements. First, there is the continuous cost of infrastructure, which includes the replacement of equipment (such as new transceivers), the lease costs of equipment (such as servers) and energy consumption. Second, Customer Relationship Management requires a helpdesk and marketing efforts. Third, the assets (equipment and software) must be maintained. Finally, the control centre needs to be up and running. Lease and maintenance costs are estimated as a fixed percentage of new costs (resp. 20% en 15%). The energy consumption is calculated by using the bill of material and equipment estimates. Marketing is considered to be a government campaign, which starts initially at 250.000 EUR, but decreases over time, proportionally to the number of non-users. For the helpdesk and the control centre, we estimate a number of employees based on the applications. Their wages are the main cost driver, but we take a general overhead of 30% into account.

A break-down phase doesn't seem relevant in this case.

4. RESULTS

Figure 4 shows the impact of the adoption scenario and the selected medium to long-range unicast technology. Both safety and recreational applications are rolled out. The other scenario building blocks were set to identical and logic values. Their impact is discussed further.

The cost figures shown are the Net Present Cost of the cash flows of the roll-out during the first 10 years. All further cost references are calculated in this fashion. A discount rate of 15% was used, because of the high risks associated with the project. As a reference, the Belgian highways are about 1.747km long.

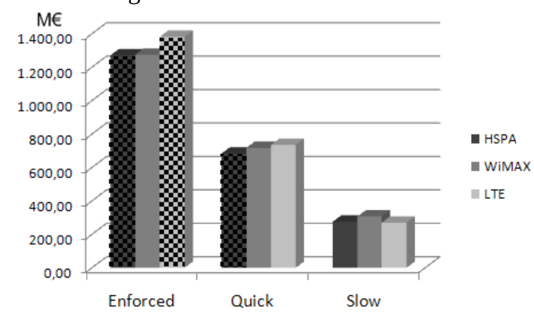


Figure 4: Cost for different adoption and technology scenarios, in case both safety and recreational applications are rolled out. Checked bars are invalid options.

Our model takes the maximum number of connections per base station and the connection duration versus connection set-up time into account. In some cases, it is technically not feasible to realize a roll-out scenario, because these constraints are broken. This is also indicated in the graph. This can be avoided by lowering some roll-out requirements (i.e. not equipping parts of the highway to handle two-way traffic jams).

In all these cases, no broadcast technology was used. The use of DVB-T causes a cost increase of about 47%. This can be explained by the cost in the OBU for the extra receiver, as well as data subscription costs for each user. MBMS in combination with HSPA realizes a cost reduction of only 1,5%, but was not used in the graph for a better comparison. The cost reduction is to be expected, because the hardware equipment is identical. However, using the broadcast functionality frees up some bandwidth of the unicast functionality, which translates in a few base stations less.

Let's investigate the impact of changing the application selection. Opting for only safety or recreational applications leads to a very similar cost, which is only about 10% to 20% smaller than combining both. Increasing the use of recreational applications to 20% is not a valid option, unless adoption occurs slowly and WiMAX is used.

We have also assumed a four-phased roll-out and a balanced mix of traffic situations. In comparison, a full roll-out from day one would cost about 68% more.

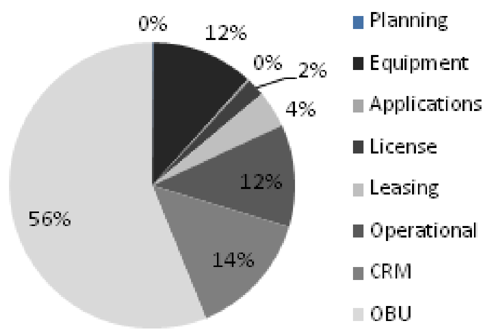


Figure 5: Break-up of total costs after 10 years for WiMAX scenario with voluntary, fast adoption.

The cost impact of re-using existing equipment is actually limited. When we compare a WiMAX roll-out with or without re-use of existing base stations and backhauling connection, the total cost is only 8% higher. We do note that a lease fee of 10% to 15% may be conservative.

In the rest of the discussion, we work with the scenario of mobile WiMAX and a fast, voluntary adoption. We will now investigate the cost break-up more in detail (Figure 5). The biggest cost by far is the OBU. This can easily be explained, considering it is a cost of a few hundred euros per user. The second most important cost is CRM, which we have calculated as a variable cost that increases in line with the total number of customers. Different implementations may avoid or lower this effect, e.g. by charging a tariff for calling the helpdesk.

A more practical way of looking at the result, is calculating the cost per user. The direct cost is the OBU, which was already discussed. If we take the sum of all non-OBU costs over the course of 10 years and divide this by the number of customer-years, we obtain a cost of about 48EUR per customer per year. Assuming an OBU also lasts for 10 years, the total cost is about 80EUR per customer per year. There is a big impact from the adoption rate. Enforced adoption only lowers this cost to about 65EUR, while a slow, voluntary adoption increases it to 250EUR.

5. SUMMARY & CONCLUSIONS

In this paper, we've presented a model that calculates the costs of rolling out an ITS on highways with a set of cooperative applications. Our approach is unique, because we've considered the joint roll-out of many applications, including recreational, and also took existing equipment into account.

The combination of applications over the same infrastructure has little impact on the total costs, as long as the bandwidth requirements remain moderate. This can be a great way to increase the value of the system, without increasing the costs. The re-use of equipment had a limited, but positive effect.

The results also indicate that especially HSPA is not a future proof technology for this type of applications and that broadcast technologies offer limited economical benefits. In its current form, LTE does not offer improved functionality over mobile WiMAX. The model indicates it is more expensive, but reliable cost estimations for LTE equipment are hard to come by and, as LTE is also much more recent, cost reductions can be expected in the upcoming years. It is also apparent that considerable savings can be made by fine-tuning the roll-out.

While it is not our intention to deliver a detailed planning and exact cost calculation tool, the absolute figures are nevertheless a good indication of the order of magnitude of the costs. The high figures can be relevant when considering the role of the government or private parties in the roll-out of an ITS.

The great impact of the OBU and CRM in the total cost makes these obvious candidates for future cost reduction investigations. Finally, we noticed that the adoption rate has a big impact on the cost, while it was very hard to gather reliable information on this input parameter. This suggests that this, too, should be the focus of further study.

ACKNOWLEDGEMENTS

The authors would like to thank Koen Casier, Didier Colle, Margot Deruyck, Maarten De Groote and Bart Lannoo for their contribution to this paper. They would also like to thank the Flemish Policy Research Centre Mobility Works, track Traffic Safety for partially funding this research.

REFERENCES

[1] S. Kang, D. Gillen, 1999, *Assessing the Benefits and Costs of Intelligent Transportation Systems: Ramp Meters,* California PATH Research Report

[2] D. Levinson, D. Gillen, E. Chang, 1999, *Assessing the Benefits and Costs of Intelligent Transportation Systems: The Value of Advanced*

Traveler Information Systems, California PATH Research Report

[3] http://www.its.dot.gov/index.htm

[4] RACC Automobile Club, 2007, *Stakeholder utility, data privacy and usability analysis and recommendations for operational guarantees and system safeguards: Europe*, Public deliverable D.DEPN.4.1 of the FP7 CVIS project

[5] H. Baum, T. Geissler, U. Westerkamp, C. Vitale, 2008, *Cost-Benefit Analyses for stand-alone and co-operative Intel-ligent Vehicle Safety Systems*, Public Deliverable D6 of the FP6 eIMPACT project

[6] ETSI, 2009, *Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Definitions*, Standard No. TR 102 638

[7] Car 2 Car Communication Consortium, *C2C-CC Manifesto*, 2007

[8] M. Bechler, H. Berninger, T. Biehle, T.M. Bohnert, R. Bossom et al., *European ITS Communication Architecture*, deliverable of the COMeSafety project

[9] R.A. Uzcátegui, G. Acosta-Marum, *WAVE: A Tutorial*, IEEE Communications Magazine, pp. 126-133, May 2009

[10] W. Vandenberghe, P. Leroux, I. Moerman, F. De Turck, P. Demeester, *Design of a Scalable ITS architecture based on IP Datacast over DVB-H/SH*, 8th International Conference on ITS Tele-communications (ITST), Phuket, Thailand, 2008.

[11] Global UMTS and HSPA operator status, available http://www.3gamericas.org/documents/Global_3G_Status_Update.pdf, 9/2009, accessed 6/2010

[12] W. Vandenberghe, D. Carels, I. Moerman, P. Demeester, J. Bergs, E. Van de Velde et al., *Road Tolling 2^E2 impact study*, deliverible of the IBBT NextGenITS project, 2009

BEST PRACTICE FOR THE GSM/GPRS/UMTS NETWORK IMPLEMENTATION INSIDE TUNNELS OF HIGH-SPEED LINES

Case Studies: Sections Madrid - Lleida – Tarragona – Barcelona – Figueres

Joaquín Fernández de Arcaya Martínez

IDOM, C/ Lehendakari Aguirre nº 3, Bilbao, Spain

jfa@idom.com

Keywords: radio communication, tunnel, high-speed line, GSM-R, GSM, GPRS, UMTS, mobile, power supply.

Abstract: This paper which has been prepared for the 49th Federation of Telecommunications Engineers of the European Union (FITCE) Congress (September 2010), analyses the requirements for the radio communication networks which should be under operation inside the tunnels of high-speed railway lines: GSM-R system which supports communications for railway traffic management (trains, maintenance brigades and signalling), GSM/GPRS/UMTS systems which supports radio mobile telephone for railway passenger and TETRA system which supports mobile communications for emergency agencies. High-speed railway infrastructure specific characteristics are described: power supply based on high-voltage railway electrification system and redundant power supply for signalling proposes. On the other hand, the application of a well pre-studied design for GSM/GPRS/UMTS radio mobile telephone installation inside tunnels of high-speed railways is outlined: possible coordination among the mobile telephone public operators, required coordination between the mobile telephone public operators and the railway infrastructure (GSM-R, fixed communications and power supply projects). Some case studies based on different strategies considered on several sections in high-speed lines are presented. Finally, TETRA implementation inside the large tunnels of high-speed lines is recommended.

1. RATIONALE

High-speed railways designed for passenger traffic are under construction all over the world to connect large cities with trains up to 350 km/h. Tunnel construction results inevitable both to through mountains and in the closed trams to the main stations in populated cities.

The mobile communication systems used in high-speed lines require specific installations to extend its coverage inside the tunnels. Therefore, it is interesting to identify the best practices for the implementation of the GSM/GPRS/UMTS network.

In general, the mobile telephone public network is independent of the radio railway installation (GSM-R). Sometimes mobile telephone public networks have not been considered as railway installations during railway line design and construction. Therefore, sometimes mobile telephone public networks which support passenger communications are based on out-track base station and tunnels result not adequate network coverage.

2. MOBILE COMMUNICATION SYSTEMS INSIDE TUNNELS

There are three different separate mobile communication systems which should be under operation inside tunnels of high-speed railway lines: GSM-R used for the voice and data private railway communications, GSM/GPRS/UMTS used for travellers mobile communications and TETRA used by governmental emergency management agencies.

2.1 GSM-R System for Railway Communications

GSM-R technology is a European standard for private railway voice and data radio communications also used in non-European countries as Algeria, China, Korea and India. On the first hand, GSM-R supports the voice communications for trains and maintenance brigades with the traffic control centres. On the other hand, GSM-R supports the radio communication channel for ETCS/ERTMS level 2 signalling system. High-speed railway lines in Europe operating ERTMS level 2 require a GSM-R network under operation as well other high-speed lines outside Europe.

2.1.1 Technical Requirements

The EIRENE recommended value for GSM-R network planning for ETCS level 2 high-speed lines above 280 km/h is coverage probability of 95% on each 100 m section based on a coverage level of -92 dBm defined as the field strength at the antenna on the roof of a train. Additionally, the handover success rate should be at last 95.5% over train routes under design load conditions. To avoid the necessity for large cell overlaps to accommodate high-speed train operations, optimisation of the handover process for such trains is considered necessary. Suitable algorithms are tested and refined as necessary during the trials process. The UIC frequency band for GSM-R system is 876 – 880 MHz (mobile station transmit - uplink) and 921 – 925 MHz (base station transmit – downlink). The carrier frequency is designated by the absolute radio frequency channel number. For carriers in the UIC frequency band the following convention shall be used, where Fl is the frequency value in the lower band and Fu is the corresponding frequency value in the upper band (frequencies in MHz) [GSM-R Operators Group, 2006. UIC Project EIRENE. System Requirements Specification. UIC]:

$$Fl(n) = 890 + 0{,}2*(n-1024);\ 955 \le n \le 973 \qquad (1)$$

$$Fu(n) = Fl(n) + 45\ \text{MHz} \qquad (2)$$

Popular Republic of China Ministry of Railways (MOR) selected GSM-R technology as the main mobile communication system for both high-speed and conventional new railways line sections in China. On China railways, GSM-R supports the radio communication channel for CTCS signalling system.

2.1.2 GSM-R Network Operators and Users

GSM-R network operators in the world are organizations dependent of the railway infrastructure administrator. On the other hand, the railway companies operating trains in the high speed lines are GSM-R user systems.

Railway companies which are railway operator, facilitate fast high-speed test trains used for GSM-R network operator to on track confirm the GSM-R quality of service test.

Table 2: Some European GSM-R network operator and users.

Country	GSM-R network operator	GSM-R user Railway operator
France	RFF	SNCF
Germany	DB Netz	DB Ag
Italy	RFI	Trenitalia
Poland	PLK SA	PKP SA
Spain	ADIF	Renfe

2.2 GSM/GPRS/UMTS

GSM (2G) is the standard for mobile telephone system in Europe and other countries in the world. GPRS (2.5G) is an extension of GSM allows data rates of 56 - 114 Kb/s. UMTS is the standard for 3G.

Although GSM/GPRS/UMTS are mobile communication systems independent from the railway operation, railway administrators consider these services of interest for passengers of high-speed lines. Consequently, railway administrators started to facilitate coverage works, both on surface and tunnel trams of high-speed lines, to mobile telephone public operators. Test trains used by GSM-R quality of service system test are used for GSM/GPRS/UMTS systems test.

2.2.1 Technical Requirements

The frequency bands used for GSM in Europe are: E-GSM 900 and P-GSM-900 bands are 880 – 915 MHz (uplink) and 925 – 960 MHz (downlink), DCS-1800 bands are 1,710 – 1,785 (uplink) and 1,805 – 1,880 (downlink). Therefore, GSM-R and GSM frequencies band are closed. Nevertheless, the coverage level required is +22 dB for GSM due coverage is required inside the trains which are Faraday cages.

The UMTS (3G) frequency band used in European countries is 2.1 GHz. Additionally, the 2.5 GHz frequency band is expected to be used for UMTS operation. Therefore, UMTS frequencies are quite distance to GSM-R and GSM frequencies. Consequently, UMTS requirements are more restrictive for network planning.

2.2.2 Mobile Telephone Public Operators in Spain

The three main mobile telephone public operator infrastructures in Spain are France Telecom España, S.A. (Orange), Vodafone España, S.A.U. and Telefónica Móviles España, S.A.U. (TME). These three companies participate in the mobile projects for tunnels in high-speed lines. On the other hand, Xfera Móviles, S.A. (Xfera) which is the forth 3G mobile operator in Spain, is not present in the mobile projects for tunnels of high-speed lines. Xfera signed a national roaming agreement with Vodafone before start the service in 2006. Additionally, in 2008, Xfera signed a national roaming agreement with TME. Therefore, the Vodafone and TME networks support Xfera users inside tunnels of high speed lines.

In Spain, the competence among the mobile telephone networks exists due law requirement. Although some equipment at the site and inside tunnel installation is common for all the operators, it is required that each network should be independent.

2.3 TETRA

TETRA is an evolution of analogue trunking MPT1327. TETRA is an ETSI European standard based on TDMA digital communications designed for group mobile communication operation. Tetra is used by government agencies and emergency services (medical services, fire brigades and police) in the reserved 380 – 385 and 390 – 395 MHz frequency bands. TETRA is replacing both PMR and other digital trunking systems implemented in Europe for government agencies and emergency services as TETRAPOL. Although TETRA works at speeds up to 400 km/h, TETRA coverage inside tunnels of high-speed lines is oriented, where it exits, only for emergency situations inside large tunnels.

2.3.1 TETRA Implementation in Very Large Railway Tunnels

The most large railway tunnels in the world are: Seikan tunnel (Sinkansen - Japan, 53.8 km), the Channel Tunnel (high-speed Eurostar passenger trains at 300 km/h - France – United Kingdom, 50.5 km), The Lötschberb Base Tunnel (tilting passenger trains at 250 km/h - Switzerland, 34.5 km) and the railway tunnels of Guadarrama (high-speed line from Madrid to Valladolid with trains over 300 km/h - Spain, 28 km).

There are TETRA network installations for coverage inside the Channel Tunnel Rail Link and large railway tunnels of Guadarrama. Both Channel Tunnel and Guadarrama tunnels design consist of two parallel tunnels which are connected each 250 m, separated 30 m each form the other.

2.3.2 TETRA Coverage Extension in Railway Tunnels Based on DMOs

TETRA permits direct communications in areas of bad coverage using the direct mode operation facility: gateway or repeater. In this case, terminals operate as repeaters for TETRA signal. Therefore, the extension of TETRA network coverage has not been considered in short railways tunnels.

2.3.3 TETRA Implementation in Underground Rail Systems

TETRA implementation in underground rail systems (Metro Bilbao, Metro Madrid, London Underground, Paris Metro, etc.) is not referenced in this paper.

2.3.4 European Regulations Considering Radio Coverage for TETRA

The Directive 2004/54/CE (*Real Decreto* 635/2006 in Spain) requires radio coverage inside large road tunnels for emergency services supported by tunnel infrastructure administrator. Therefore, depending on the radio system, TETRA installation would be considered in road tunnels.

On the other hand, the Commission Decision of 20 December 2007 (2008/163/CE) indicates for railways: “Radio continuity shall be provided for permitting the rescue services to communicate with their on-site command facilities. The system shall allow the rescue services to use their own communication equipment.”.

3 REQUIRED INFRASTRUCTURES FOR GSM/GPRS/UMTS EXTENSION INSIDE TUNNELS

GSM/GPRS/UMTS technologies require specific installations inside the tunnels of high speed lines: antennas, active radio FOR equipment, a power supply system and optical fibre installation. The installation is based on antennas fixed to the tunnel at concrete positions instead of radiating leaky coaxial cables along the tunnel. First, antennas are located at the vicinity of the tunnel entrance or exit. On the other hand, the installation of additional antennas is required in large tunnels (each 600 - 800 m). The antennas are connected to local FOR equipments which consumes about 200 W measured during operation (400 W maximum power indicated by the manufacture). In case of very large tunnels, specific base stations are installed inside tunnels usually in some of the connecting galleries. The infrastructures required to connect the active radio communication equipments inside the tunnels (FOR) with the base station outside the tunnel are quite simple: electrical power supply and optical fibre cables. The following figure shows an adequate power supply (operators site and tunnel box power supply are connected both to the high-voltage railway electrification system and to the redundant power supply for signalling proposes) installation.

3.1 Railway Infrastructures for Power Supply

High-speed lines power supply to trains is based on high-voltage electrification lines, typically 1x25 kVca or 2x25 kVca, from the electrical substations located along the track (reference distance 60 km, input voltage 200 kVca, 400 kVca). The catenary, technical buildings and radio base stations (GSM-R or GSM/GPRS/UMTS) are connected to the same electrification line. On the other hand, there is a 600 Vca (case of Spanish high-speed lines) continuously powered electrical line from the technical buildings for signalling proposes: signalling remote insta-llations, tunnel entrance or exit equipment buildings (including the tunnel illumination), GSM-R sites and early alert and safety systems (meteorological stations including anemometers and snowfall detectors, wind speed meters, landslide and rock falls detecting systems, etc.). Due catenary requires to be cut off for maintenance works, high-speed technical buildings dispose of diesel generators and batteries which support UPS system for both inside equipment and consumers along the track which are connected to the continuously powered electrical line. Consequently, there are two power supply networks available on high-speed lines: power supply based on high voltage railway electrification line system (25,000 Vca) and redundant power supply for signalling proposes (600 Vca).

3.1.1 High-Voltage Railway Electrification Line System

A transformer connected to the high-voltage railway electrification line system (25,000 Vac) at the entrance or exit of the tunnel generates 230 Vac to power supply all the consumers of the tunnel entrance or exit equipment building or GSM-R base station. Therefore, air conditioning equipments, wall sockets for maintenance proposes or non-emergency tunnel illumination only receive power supply from the high-voltage line.

3.1.2 Redundant Power Supply for Signalling Proposes

The redundant power supply is continuously powered from one of the two technical buildings that limit the electrical section.

A transformer connected to the redundant power supply for signalling proposes (typically 600 Vca in Spanish high-speed lines) at the entrance or exit at the tunnel generates 230 Vac to power supply only the critical consumers of the tunnel entrance or exit equipment building. Some consumers both power supplied from the high-voltage line and redundant

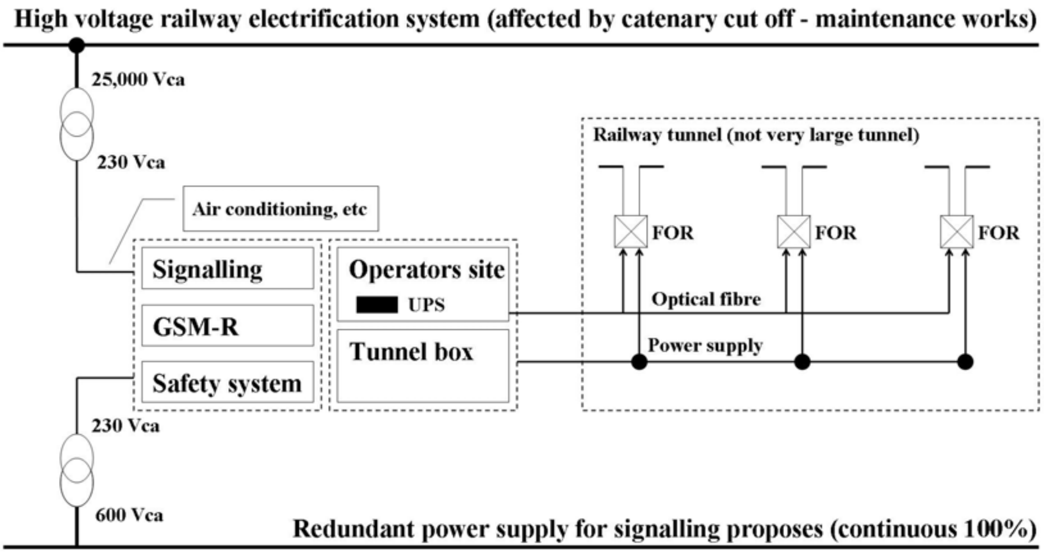


Figure 1: High-speed power supply system for consumers.

power supply are signalling systems, emergency tunnel illumination, etc.

3.2 Dedicated Optical Fibre

Typically, one or two dedicated optical fibres from the mobile telephone public operator building where GSM/UMTS base stations are located to each FOR equipment inside the tunnel are required. Additional dedicated optical fibres should be reserved for maintenance proposes or backup proposes.

In some cases, optical fibre cable which supports GSM-R FOR connections supports GSM/UMTS FOR connections.

4 BEST PRACTICES FOR SUPPORTING GSM/GPRS/ UMTS INFRASTRUCTURES

The best practices for infrastructures which support GSM/GPRS/UMTS networks inside tunnels of high-speed lines are: collaboration among the mobile telephone public operators in tunnel coverage project observing competence rules, inside tunnel active equipment based on FOR equipment and antennas shared and installed by the mobile telephone public operators, power supply and optical fibre installed by the tunnel infrastructure administrator (GSM-R and fixed telecommunications projects), connection both to the high-voltage railway electrification system and to the redundant power supply for signalling proposes.

5 CASE STUDIES

The four case studies following presented correspond to the evolution in the design of mobile telephone public operators incorporation to different sections of the high-speed line Madrid – Barcelona.

5.1 Case Study. Section Madrid - Lleida

The commercial service in the section Madrid – Lleida started without GSM/GPRS/UMTS coverage inside the tunnels.

After a time, the mobile telephone public operators formed an association in order to optimize the works for coverage extension inside the tunnels. The infrastructure administrator supported the installation: the association of mobile telephone public operators installed a transformer connected to the high-voltage railway electrification to power supply the installation. On the other hand, mobile telephone public operators incorporated UPS equipment to prevent the effect of catenary cut off for tunnel active equipment. The UPS was sized only for no-radio traffic during the night.

5.2 Case Study. Section Lleida – Tarragona. Support.

The infrastructure administration actively supported the mobile telephone public operators in order to dispose GMS/GPRS/UMTS coverage inside the tunnels as soon as possible the commercial service in the section Lleida – Tarragona started. A position is reserved in the site for future Xfera incorporation.

Both the power supply from the high-voltage railway electrification and the optical fibre infrastructure for FOR equipment inside the tunnels were facilitated by the infrastructure administration.

5.3 Case Study. Section Tarragona – Barcelona. Common UMTS rack

The maximum possible coordination among the mobile telephone public operators was obtained in the tunnels from El Vendrell to Martorell in the section Tarragona - Barcelona: the rack of UMTS equipment supports all the different radio equipment installed by each mobile telephone public operator. No more integration is possible due law requires independent networks for competence among companies: three GSM and three UMTS equipments are required.

5.4 Case Study. Section Barcelona – Figueres. Power supply

This section continues under construction.

The mobile telephone public operators expect that the tunnel active equipment will be supported both by the high-voltage railway electrification and the redundant power supply for signalling proposes.

6 CONCLUSIONS

6.1 GSM-R Design Conditioned by the GSM/GPRS/UMTS

The GSM-R system design for a high-speed line should consider both the GSM-R and mobile telephone public operators installation at the same sites (common site allocation and antennas supported by the same telecommunication tower on the site): First, a project coordination should exists to identify the best sites both for GSM-R and mobile telephone public operators. Subsequently, coordination should exist during works execution.

If coordination exits then a site would be placed closed to the entrance or exit of each tunnel of the high-speed line. Additionally, coordination is required for different antennas situation inside the tunnel. Finally, coordination is required to avoid interferences between GSM-R and GSM/UMTS radio equipments.

6.2 Infrastructure Administrator Collaboration

Power supply and optical fibre for FOR equipment inside the tunnel should be facilitated by the infrastructure administrator (GSM-R and fixed communication projects). Consequently, an easy project organization and simple work for mobile telephone public operator will de achieved.

6.3 Power Supply for Tunnel Equipment

Continuous power supply is recommended to enlarge the equipment life. Therefore, radio equipment inside tunnels of high-speed lines should be continuously powered. Therefore both a connection with the high-voltage electrification system and the redundant power supply for signalling proposes should be considered during the design phase of both mobile telephone public operators system and the energy system.

6.4 Railway Company Collaboration

An absolute necessary coordination between the mobile telephone public operators and the railway company for test of GSM/GPRS/UMTS coverage on board trains is required.

6.5 On Board Train Equipment

In the future, technical solutions based on leaky coaxial cables and GSM/GPRS/UMTS radio repeaters on board trains of high-speed lines will be analysed to improve the coverage inside the trains which are Faraday cages. These installations will require the coordination among the railway operators and mobile telephone public operators.

6.6 TETRA Coverage

TETRA coverage for government agencies and emergency services should be considered in large tunnel of high-speed lines design. If an emergency happens in a not TETRA covered tunnel then TETRA coverage should be extended based on DMO facilities until specific regulations are approved.

REFERENCES

GSM-R Operators Group, 2006. UIC Project EIRENE. System Requirements Specification. UIC.

CE Commission Directive 2004/54/CE.

Spanish *Real Decreto* 635/2006.

CE Commission Decision of 20 December 2007 (2008/163/CE).

ABREVIATIONS

ADIF: *Administrador de Infraestructuras Ferroviarias* (Spanish railway infrastructure administrator).
CTCS: Chinese Train Control System.
DMO: Direct Mode Operation.
EIRENE: European Integrated Railway Radio Enhanced Network.
ERTMS: European Railway Traffic Management System.
ETCS: European Train Control System.
FOR: Fibre optical repeater.
GSM: Global System for Mobile communications.
GSM-R: Global System for Mobile Communications – Railway.
GPRS: General Packet Radio Service.
MOR: Popular Republic of China Ministry of Railways.
PKP SA: *Polskie Koleje Państwowe SA* (Polish State Railways).
PLK SA: *Polskie Linie Kolejowe SA* (Polish railway infrastructure administrator).
RFF: *Réseau Ferré de France* (French railway infrastructure administrator).
RTI: *Rete Ferroviaria Italiana* (Italian railway infrastructure administrator).
TDMA: Time Division Multiple Access.
TETRA: Trans European Trunked Radio.
UIC: *Union Internationale des Chemins de fer* (International Union of Railways).
UMTS: Universal Mobile Telecommunications System.
UPS: Uninterruptible Power Supply.

A PROPOSAL FOR TRAFFIC GUIDANCE AND SIGNALLING SYSTEM BASED ON RADIOFREQUENCY BEACONS

FITCE 2010 Congress

Luis Fernández Navarro
Senior Telecomunication Engineer
luis.fernandez@coit.es

Keywords: Traffic guidance, signalling system, radiofrequency beacons, IPv6, speed limits, way outs, networking system, intelligent speed adaptation (ISA), RIPE, ITU, ETSI.

Abstract: As countries acquire higher levels of development, the traffic management becomes more complex and expensive. As a consequence, the development of an automatic traffic guidance system represents a solution for traffic management, reducing the cost in human lifes and allowing research innovation and society economical growth. In these lines a proposal on both protocol and radio aspects to implement the first steps for an automatic traffic guidance and signalling system based on radiofrequency beacons is presented.

1 INTRODUCTION

Terrestrial traffic management represents an important task for countries authorities, as an appropriate goods and workers transport net is related with the productivity and economical growth. Moreover, the treatment of human and material crashes consequences causes a high economical cost.

Traffic signals present in roads are not adaptive to changes in the environmental conditions (weather [*17,6% of traffic accidents occurs on bad weather days*[1]], traffic jams, collisions, daylight or night), so, the information provided by these signals may be inadequate or even damaging according to the traffic conditions.

A high percentage of crashes causes can be grouped into **Inadequate speed** and **Way outs**:

Table 1: Crashes causes percentages (some EU countries).

Country	Netherland[2]	U.K.[3]	Spain[4]
Speed	20,00%	33,33%	49,97%
Way outs	33,33%	-	41,70%

There are several Intelligent Speed Adaptation (ISA) solutions[5-6-7], but usually they use the static road signals information: *speed maps* (use GPS or similar), or *image processing* (of posted signals).

The traffic system has many similarities with a networking system: vehicles **vs** data packets, roads **vs** transmission lines, stop and give way and others signals **vs** QoS rules that prioritize some queues, direction signals **vs** routing tables, distances **vs** metrics of the routings protocols, countries and cities **vs** IP sub netting and postal address **vs** IP address.

This document proposes to apply the well known networking technologies to traffic management, specifically, in order to codify and manage properly destinations, indication signals and traffic system identifiers, the use of IPv6 is proposed.

The proposed system faces the detailed problems using a radiofrequency signalling system for traffic guidance with the following advantages:

- Establish an **IPv6 addressing space** correlated with postal addresses.
- Develop a **dynamic signalling system**, adaptable to environmental conditions.
- Develop a **radiofrequency beacons guidance system** on the road lanes, avoiding way outs.

This system will increase the safety of roads, will reduce the economical cost of crashes and could be an impulse to the economy (auto motion, traffic signalling, mobile, chips, delivery industries…).

This work analyses the requirements, focusing in the described advantages. First, a correlation between IPv6 and postal address spaces is established. Second, a design of the required protocol to develop a dynamic signalling system is done. Last, the radio aspects for developing a traffic guidance system are analysed.

2 IPv6 AND POSTAL ADDRESSES

The proposed method asks RIPE an IPv6 allocation to assign a global unicast IPv6 address to each postal address on each building of each street. The assignation of the IPv6 address must be done considering the subnetting rules, so, each street must contains the summary of the buildings IPv6 address. The summary of streets addressing space is assigned to cities, and the same for regions, countries…

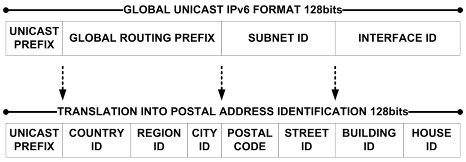


Figure 1: IPv6 – Postal address translation example.

It will be necessary to determine each section width in bits and develop the appropriate DNS service to translate the postal address into the IPv6 addressing space.

Develop this correlation might translate into economical benefits for DNS service companies and for delivery companies because each postal address could be translated into a univocal IPv6 address, avoiding confusions.

3 PROTOCOL OF DYNAMIC SIGNALLING SYSTEM

3.1 Identification of the Reference Signalling Information

As seen in figure 3, one radio beacon transmitter is able to cover the two road lanes, therefore, the same radio beacon is transmitting the signalling information for both road directions. In order to distinguish the reference information each radio beacon has as unique identifier an IPv6 address, so each IdRB (Id Radio Beacon) structure is:

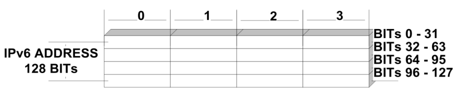


Figure 2: IdRB structure is an IPv6 address.

Moreover, it knows the neighbours identifiers and transmit the sequence of identifiers for each road direction. In that case, the vehicle, which knows the radio beacon identifiers which just have covered it, is able to extract the correct information.

In order to explain in a better way the proposed operating method, next figure shows a vehicle which is moving in direction to a radio beacon with identifier IdRB005:

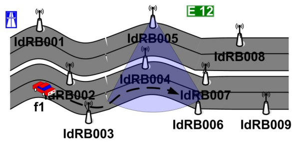


Figure 3: A radio beacon covers both road lines.

IdRB005 is transmitting this information:

- Direction A: IdRB002 - IdRB003 - IdRB004 - IdRB005 - Signalling Information A.
- Direction B: IdRB005 - IdRB006 - IdRB007 - IdRB008 - Signalling Information B.

When the vehicle arrives to the coverage area of IdRB005 knows that it just has been crossed the coverage areas of IdRB002, IdRB003 and IdRB004, so its direction is A and it can decode the reference signalling information payload (SIP) A.

The information stack is sent alternating in time each road direction information:

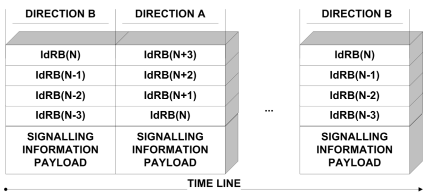


Figure 4: IdRB and SIP stack time line structure.

3.2 A Signalling Information Payload (SIP) Protocol

The proposed system must allow inform about traffic signals with the following features:

- Several traffic signals on the same point or coverage area.
- Complementary information like: **Range of applicability** (*Space range aimed*, I.e., reversal way in 500m advertising. *Time range aimed*, I.e., a speed limit until sunrise or sunset), **Mandatory character** (For instance a stop or a give way, and must be notified to drivers by the proposed system. Other, like informational signals, does not have that meaning and theirs advertisements could be

disabled by drivers). **Indication signals** (directions, city names, exit numbers…).

Consequently, each SIP byte is assigned the meaning contained in figure 5:

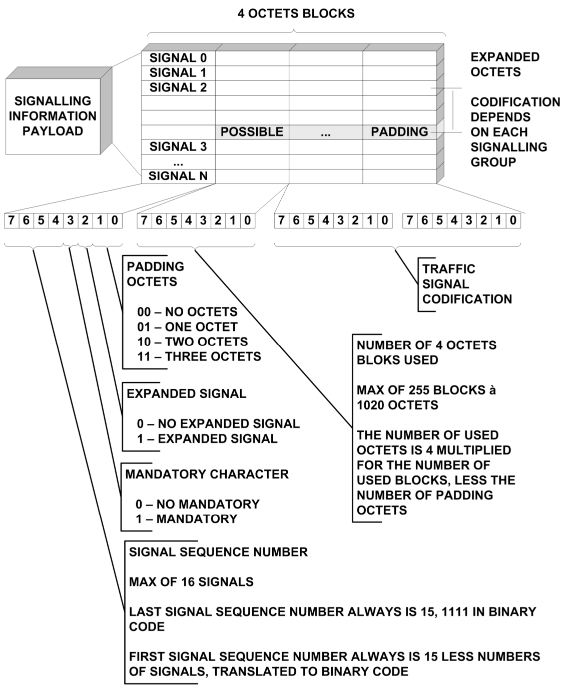


Figure 5: SIP structure.

The protocol proposed uses the four most significant bits (4-7) of the most significant octet to identify the signal sequence number. It is supposed a maximum of 16 signals to advice per direction on each radio beacon. Always the last signal will have the identifier 1111, so it becomes necessary to assigns the identifiers in decreasing order and then stack the signals information in increasing order. Bit 3 informs about the mandatory character of the traffic signal and bit 2 informs about the constitution of the following octets. If bit 2 has a value of 0 the signal doesn't require others octets blocks the number of blocks used will be one (00000001) and the signal will be codify in the latest two octets of the block. If bit 2 has a value of 1 it will mean that the signal information requires more than one block, the number of used blocks will be on the second octet. The second octet value 00000000 will be reserved, other values will represent the number of used blocks. If the signal information is using more than one block it is possible not to fill the last octet. To inform about this case there are the two less significant bits of the most significant octet: **00** No padding octets, **01** One padding octet, **10** Two padding octets, **11** Three padding octets.

It is possible to establish a common pattern for the meaning of the expanded octets, so as to supply requirements about the range of applicability (on space and time domains), directions indications and other specific information and meanings (depending on the given signal). The following bit utilization is proposed to this end:

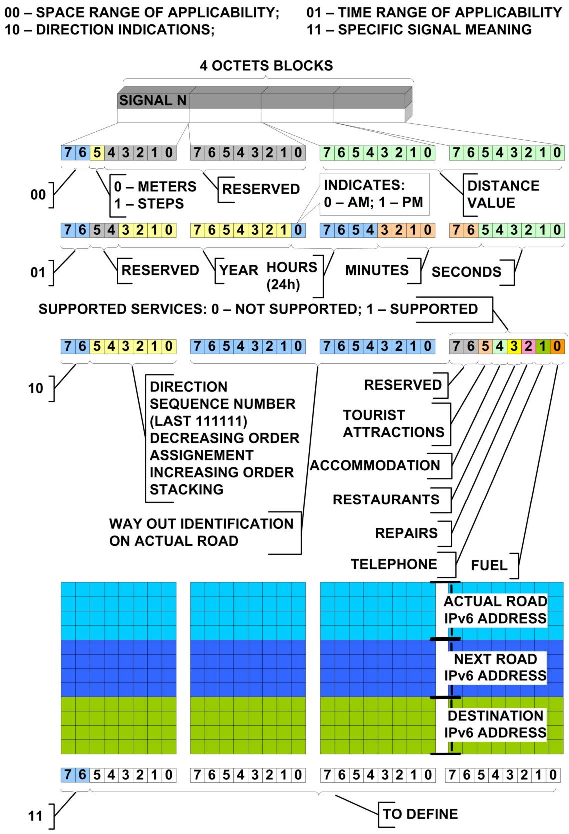


Figure 6: Expanded octets bit utilization.

As can be observed in figure 6, similar sequence numbers are used in directions indications: the value of 111111 indicates that we are in the last direction indication. Therefore, the values are assigned in decreasing order and stacked in increasing order. Moreover, there are two octets in the directions indications aimed to codify the way out identification number in the current road. Also, there is an octet to specify the supported services as petrol stations, telephone, repairs, restaurants... Finally, two bits are reserved for future use. As was detailed before, the IPv6 address of the current road, next road and destination are used to indicate a direction.

3.3 Proposal of Codification of the Traffic Signalling System

In a first level, traffic signals can be classified using both the Commonwealth Classification and the Vienna Convention Classification. In a second level, the classification method must follow the traffic regulation, so the signals can be classified in:

- **Regulatory signals:** Priority, Forbidden entry, Restricted entry, Obligation, End of forbidden or obligation.
- **Danger warning signals.**
- **Indication signals:** General indication, Lanes, Service, Orientation (*Signalling in advance, Direction, Road Identification, Localization, Confirmation, Specific use in towns*), Additional panels.
- **Works signals:** Danger, Priority and regulation, Indication.
- **Traffic lights.**
- **Other signals.**

In order to translate this classification into a binary codification system the use of two octets is proposed, the first of them to identify the type of signal and the second to identify the specific signal in the group. In the first octet, the most significant bit will be used to distinguish the Commonwealth and the Vienna Convention Classification Groups. The following three bits are used to identify the second level of classification group and the four less significant bits will acquire significance depending on the second level classification group. The second octet is used to identify the specific signal in its classification group. The proposed protocol for signals codification is detailed in figure 7.

Moreover, each country has their specific traffic signals that do not fit in the general classification, so the use of the three reserved second level bits on the first octet and the four third level bits in the same octet is proposed to identify the Country on the Vienna Convention (0110xxxx) and on the Commonwealth (1111xxxx), while the second octet will be used to identify the specific traffic signal:

Table 2: Examples of Country codification criteria.

Level 1	Country	Octect 1	Octect 0
Vienna Conv.	France	01100000	--------
	Germany	01100001	--------
	Spain	01100010	--------
Commonwealth	Ireland	11110000	--------
	U.K.	11110001	--------
	U.S.A	11110010	--------

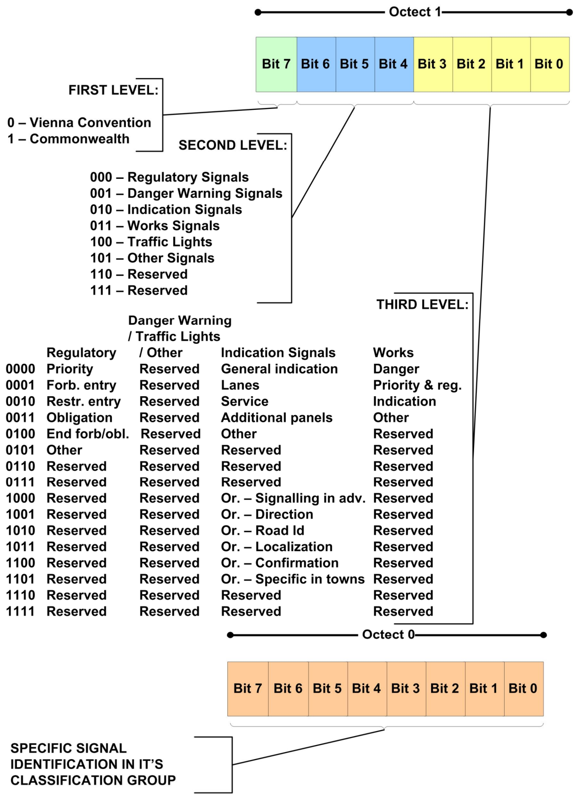


Figure 7: Signals codification.

Table 3: Examples of how traffic signals are encoded.

Signal	L1/L2/L3	Octect 1	Octect 0
	V.C./Reg/Priority	00000000	00000000
	C.W./Reg/Priority	10000000	00000000
	V.C./Reg/Priority	00000000	00000001
	C.W./Reg/Priority	10000000	00000001
	V.C./Reg/Forb.Entry	00000001	00000000
	C.W./Reg/Forb.Entry	10000001	00000000

The protocol to assign values to the octet 0 will depend on the specific signal classification group.

3.4 Dynamic Speed Limit System

In order to implement a dynamic and adaptive speed limit system it's necessary to codify the several speed limits. Some of the pre-requisites are:

- Codify the speed limit metric system (obtained from the first encoding level): *Vienna C.-Kilometres, Commonwealth-Miles*.
- Codify the no established speed limit.
- Codify the environmental conditions, at least eight causes (Requires tree bits on the octet 0): *Rain, Fog, Frost, Daylight, Night, Deficient road illumination, Traffic Jam, Collisions*.

- Codify speed values between 5 and 150 (highest speed limit in km [higher than in miles] in some EU regions) in intervals of 5. It's proposed to use the last five bits in octet 0: *Zero value is the no established speed limit - Each value represent unit interval of 5. The speed limit is the given value multiplied by 5.*

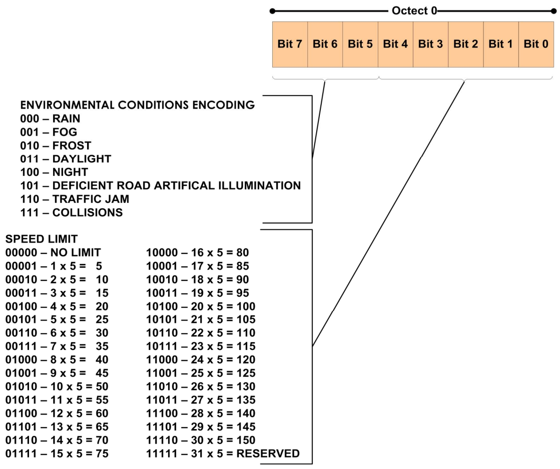


Figure 8: Speed limits codification.

4 RADIO ANALYSIS

4.1 Technical Aspects for Developing a Traffic Guidance System

For the purposes of traffic guidance, it is desirable to have a flat front wave, in order to receive a specific signal strength range in each road line. A solution is to locate the radio beacons along each road side to conform a wave front as flat as possible. The distance between radio beacons will depend on the antenna radiation patron.

Under these conditions it's possible to know the received power range that identifies each road line using the Friis Transmission Equation[8] (1):

$$P_r = P_t G_t G_r \left(\frac{\lambda}{4\pi \Re} \right)^2 \qquad (1)$$

In order to measure a Road Line received power range (RL_{rpr}) from a specific radio beacon (IdRB), we have to consider the line width (l_w), the vehicle width (v_w) and the road margins width (m_w) as is shown in figure 9. The equations of RL_{rpr} are (2) (maximum deviation to left) and (3) (maximum deviation to right):

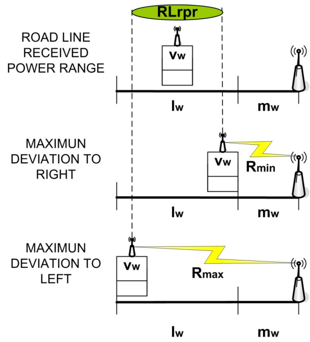


Figure 9: RL_{rpr}[IdRB] calculation.

$$RL_{rpr} \geq P_t G_t G_r \left(\frac{\lambda}{4\pi \Re_{\min}} \right)^2 \qquad (2)$$

$$RL_{rpr} \leq P_t G_t G_r \left(\frac{\lambda}{4\pi \Re_{\max}} \right)^2 \qquad (3)$$

Where R_{min} and R_{max} are calculated in equations (4) and (5) as is detailed in figure 10:

$$R_{\min} = \sqrt{\left(m_w + l_w - \frac{v_w}{2} \right)^2 + d^2} \qquad (4)$$

$$R_{\max} = \sqrt{\left(m_w + l_w + \frac{v_w}{2} \right)^2 + d^2} \qquad (5)$$

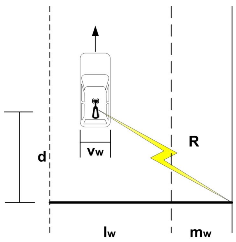


Figure 10: R calculation, d is the displacement of the vehicle over the shortest sight line to the radio beacon.

$$d_i = v_i \cdot T + \frac{1}{2} a_i T^2 \rightarrow d = \sum_{i=0}^{n} d_i \qquad (6)$$

To calculate d distance (6), the vehicle measures the points of maximum power received, and them calculates d from them as the addition of d_i measures calculated in T periods (small enough that it is valid to assume that the vehicle acceleration is constant). The method is exposed in figure 11, where the radio beacons use different frequencies in each road side:

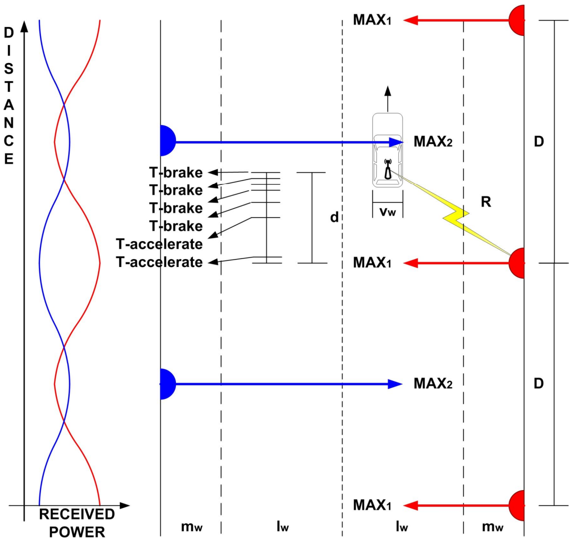


Figure 11: d calculation.

An extrapolation to the radio beacons location only in one road side can be done. The equations parameters are known or can be measured by the vehicle receptor. A zone location map may be transmitted by the radio beacons, conformed at least by each theoretical RL_{rpr}, it will be necessary to define the correspondent protocol. In fact, a road line is identified in terms of radiofrequency power by an overlapping coverage zone (OCZ), figure 12, with a specific RL_{rpr} for each radio beacon, therefore, each vehicle measures the received powers and calculates its location on road (7):

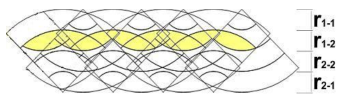


Figure 12: Road line OCZ.

$$OCZ = \sum_{n=-m}^{+m} RL_{rpr}\left[IdRB_n\right] \qquad (7)$$

The exposed equations can be extrapolated to another type of roads as highways.

4.2 Technical Aspects to Consider

The radiofrequency spectral bands that can be assigned to this service must be set by the European and International Authorities (ETSI, ITU…). The width of the radio channels will came determined by the modulation system used, the channel codification and the amount of information to transmit.

It's possible to establish some similarities with traditional mobile networks:

- Use of sectorial antennas, which allow space coverage, reducing the mutual interference (depends on the spatial distribution).
- Use of handover between transmitters.
- Use of wide spectrum techniques, as CDMA, allow a better spectrum utilization.
- AES, MD5 security protocols, to protect the system against external interferences or intentional attacks.
- MIMO techniques in order to improve the signal-noise and signal-interference rate.

In order to minimize the global system interference, it will be necessary to optimize the use of the available frequencies, the transmitted power, antennas orientation and theirs radiation pattern.

5 CONCLUSIONS

This work presented the main ideas for creating an automatic traffic guidance and signalling system based on radiofrequency beacons. The system incorporates some advantages, highlighting the establishment of an IPv6 assignation system to postal address, the development of a dynamic signalling system extended to most of traffic signals and the development of a traffic guidance system.

These advantages will allow prevent excess speed and way out accidents, will reduce the fatalities on roads, and would be traduced in an improvement of the field related industries.

The system represents an opportunity of economical growth and to export EU's technologies.

REFERENCES

[1] E.R.S.O. SafetyNet. *Annual Statistical Report 2007*
[2] European Commission, Mobility and Transport, Road Safety Observatory. *Country Profile, Netherlands.*
[3] European Commission, Mobility and Transport, Road Safety Observatory. *Country Profile, United Kingdom.*
[4] Dirección General de Tráfico, Observatorio Nacional de Seguridad Vial. *Accidentes mortales en carretera, año 2009, Spain.*
[5] Institute for Transport Studies, University of Leeds. *ISA Proyect, United Kingdom.*
[6] ETSC Speed Fact Sheet 7. *2010.*
[7] ETSC Speed Fact Sheet 4. *2009.*
[8] Hernando Rábanos, J.M. *Transmisión por radio*. 1998.

Training platform on fieldbus networks for automotive applications

Santiago Otero, Francisco Poza, Miguel A. Domínguez and Perfecto Mariño
Electronic Technology Department
University of Vigo
E.T.S.I. Industriales, Campus Lagoas-Marcosende, 36310
SPAIN
Email: jsotero@uvigo.es, fpoza@uvigo.es, mdgomez@uvigo.es, pmarino@uvigo.es

***Abstract*—The increase of the electronic equipment in the automotive market has been very important in the last years. There are amount of Electronic Control Units (ECUs) and a reliable and real-time communication between them is essential. The use of industrial communication networks or fieldbuses becomes necessary. Thus, it is important to train professionals with knowledge in the fieldbuses that are used in vehicles. This paper presents a course about fieldbuses with application in automotive industry and the laboratory implementation for the development of the practices.**

I. Introduction

Nowadays, automobiles are more and more equipped with electronic devices that make it easier to drive the vehicle and improve their security and comfort. These electronic devices are applied to functions such as Electronic Stability Program (ESP), braking help system (ABS: Antilock-Braking System), gear control, light control, climate control, seats control, navigation and guide (based on GPS: Global Positioning System and GIS: Geographic Information System), etc. These functions require the use of reliable and real time exchange of information between the different control systems and the sensors and actuators.

The ever increasing diverse population of number and complexity of electronic systems included in the automobiles, made impossible to implement this exchange of information through point-to-point links because it would suppose a disproportionate length of cable, a rising cost and production time, reliability problems and other drawbacks. Thus, it becomes necessary the use of an industrial communication network or fieldbus [1], [2], [3].

At the beginnings of 1980s the engineers of the automobile manufacturers assessed the existing fieldbuses systems for their using in vehicles. They came to the conclusion that any of these protocols fulfilled completely their requirements. It supposes the beginning of the development of new fieldbus protocols [4]. Each manufacturer has bet for a particular solution. For example, Bosch developed the CAN (Controller Area Network) protocol, Volkswagen implemented the A-BUS (Automobile Bitserielle Universal-Schnittstelle), Renault and the PSA Consortium used the VAN (Vehicle Area Network) protocol, BMW tried it with the M-BUS and Honda with the DLCS. The majority of these automobile manufacturers evolved and adopted for the general purpose communication the CAN standard [5], [6], [7], [8].

For other functionalities, such as low speed smart sensors, multimedia, high speed and safety applications, the manufacturers are adopting other protocols in the last years. For example, the Firewire (IEEE 1394), MOST (Media Oriented System Transport) [9], [10], D2B optical and D2B Smartwirex are used for high speed multimedia applications; TTP, Byteflight and FlexRay for high speed and safety applications; and LIN (Local Interconnect Network) for low speed smart sensor communication [11], [12], [13].

It is necessary to train professionals with good knowledge in these fieldbuses [14], [15]. A course about fieldbuses is described in the next sections of this paper. It is part of a Master in technologies and processes in the automotive industry which first edition was given past academic year (2008-2009) in the Industrial Engineers School of Vigo (Spain). The aim of the Master is to train recently graduates and offer new professionals to the automotive sector. The Master has 60 ECTS (European Credit Transfer System) and has duration of 10 month (October to July).

II. Design of the course

This course about fieldbuses used in the embedded automotive networks, is part of a specialization module of the Master about technologies in the automotive industry (electric and electronic systems). It is a 7 hours course and the contents are summarized in Table I.

The aim of this course is to explain the main protocols that are used currently in the automotive networks embedded in commercial vehicles [16], [17]. Thus, the core is the CAN protocol that is the most important fieldbus for the general purpose functionalities of the automobiles. There are functionalities that do not need high data rates and can be resolved with low cost nodes, and another ones that need high bandwidth (multimedia applications). Thus, the automotive communication networks have a hierarchical architecture, with networks of low data rate (up to 20 Kbps) and for multimedia application (up to 25 Mbps). The course presents the basic features of two protocols that are introducing in the vehicles for these applications that are LIN and MOST.

TABLE I
CONTENTS OF THE COURSE

Name	Topics	Hours	Type
State of the art of the automotive fieldbuses	Beginning, applications, standards	1	Theory
Design of automotive fieldbuses: the CAN protocol	Introduction to CAN, characteristics, specification CAN 2.0, history, applications, physical layer, data link layer	1	Theory
Other automotive fieldbuses: LIN and MOST	Consortiums, physical layer, frames, main features	1	Theory
Analysis tools for CAN	CANoe, CANALYZER, USB-MUX-DIAG	1	Theory
Simulation of a CAN network	Design and simulation of a CAN network using CANoe tool	1	Practice
Real CAN traffic monitoring in a simulated environment	Prototype with ACTIA ECUs, configuration of CANoe tool	1	Practice
Real CAN traffic monitoring in an vehicle	OBD connector, CANoe analyzer, identifiers	1	Practice

III. Laboratory for Training on Automotive Networks

The authors have implemented a laboratory to give an adequate training about automotive networks. This laboratory is the most interesting of the course because let to the students check the work of CAN networks, and see the frames sent by the ECUs in a real vehicle and in a prototype of automotive network. Below the practices implemented in the present course are described with the hardware and software elements that were used.

A. Practice about design software of CAN networks

The first practice introduces a software tool for designing a CAN network. The software tool chosen for this purpose is CANoe from the VECTOR company [18].

CANoe is a development, analysis and test environment for systems based in CAN networks. The CANoe development process differentiates three development stages (Figure 1):

1) The first phase defines the whole virtual system, distributing the overall functionality among different CAN network nodes, defining messages, signals, bus baud rate, etc. The user must define the behaviour of each network node using a procedural language and an event-driven model. This behavioural definition must specify what frames must be sent and received and when. Also the input and output signals to/from each network nodes are described using environment variables. The results of the simulation obtained in this phase serve to validate the design and can be used later for testing the implementation.
2) This phase covers the design and implementation of individual physical nodes, usually performed in parallel and independently, by the specific software tools from the ECUs manufacturer. The behaviour of each physical network node can be tested, using the models for the other network nodes to simulate the remainder of the bus. For this purpose a CAN interface to the real bus is required.
3) In this phase the integration of the overall system is made. All physical network nodes are connected to the CAN bus step-by-step, at the time the same virtual nodes are disconnected one-by-one from the models bus. CANoe can be used to observe the message traffic between the real network nodes on the bus, and to compare the results obtained against the specified requirements.

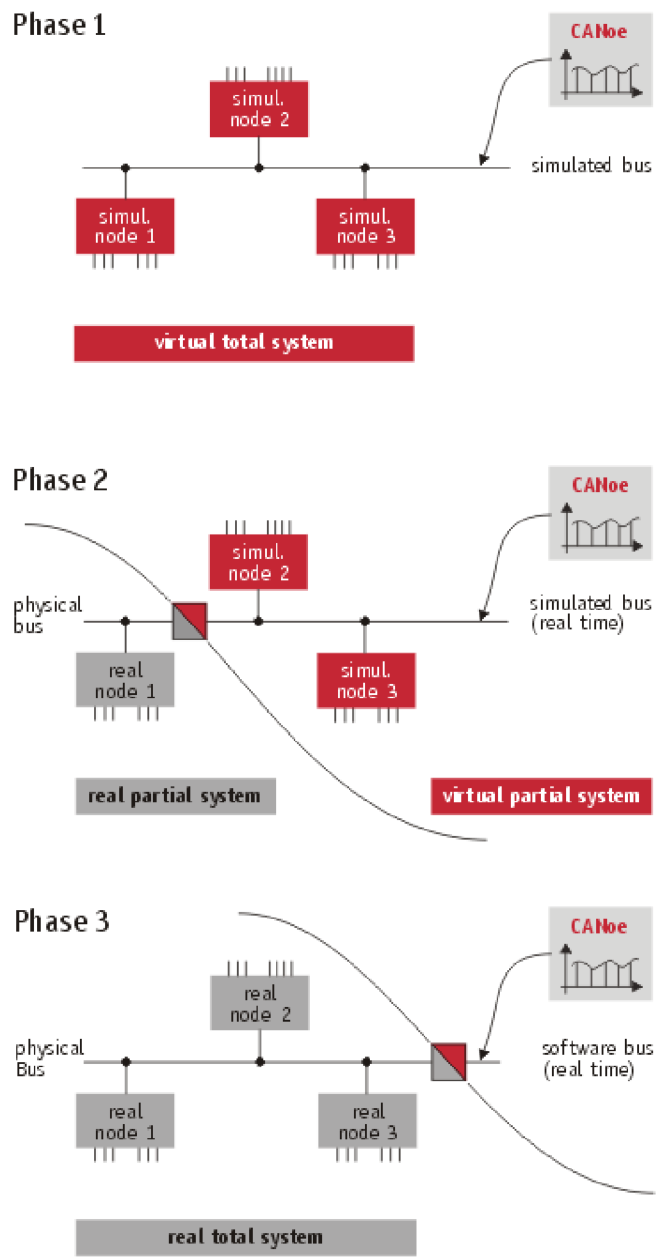


Fig. 1. Phases in CANoe development process [18]

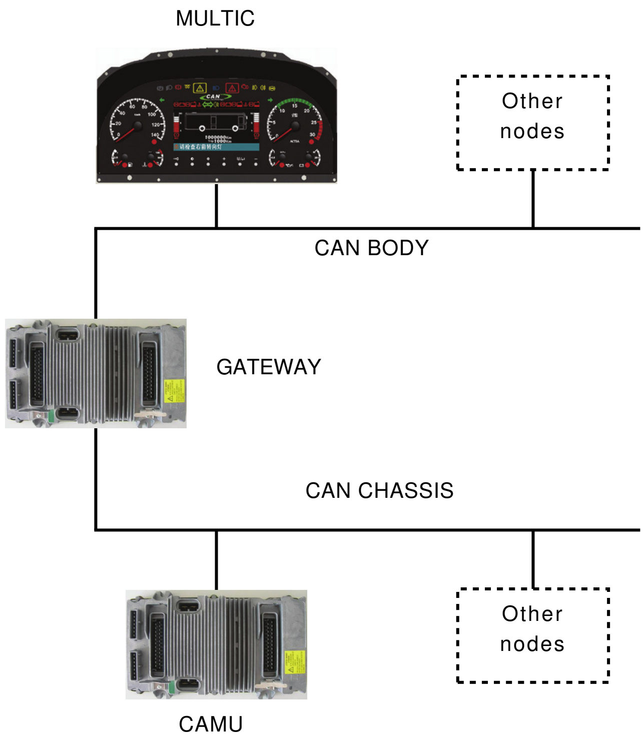


Fig. 2. CAN networks and modelled nodes

In this practice two CAN buses and three nodes have been modelled (Figure 2).

The required functionality of the system is quite simple and the operation of each network node is described below.

The CAMU (Central ACTIGRAF Management Unit) node is in the chassis network and has connected to it a position light, a DC motor and a potentiometer, simulating the operation of a gas pedal. This node receives two frames from GATEWAY node and sends one to it. The first frame received named GAT_CAMU_dig is used for the operation of the position light. The second named GAT_CAMU_ana is used for controlling the RPMs of the DC motor. The frame CAMU_GAT_ana is sent to GATEWAY with the position of the accelerator pedal.

The GATEWAY node serves as gateway between the two CAN networks and has connected a switch and another potentiometer. The switch is used to operate the position light and the potentiometer to control the RPMs of the DC motor. This node sends two frames to MULTIC node. One is GAT_MULTIC_dig to inform about the position of the switch. The other is GAT_MULTIC_ana with the same information as CAMU_GAT_ana.

Finally the MULTIC (MultiBUS Instrument Cluster) node is in the body network and is a dashboard with some gauges and lights. One light glows when the switch at GATEWAY is operated and one gauge is used to show the vehicle speed simulated with the potentiometer at CAMU. Figure 3 summarizes the frames interchanged in the system and the elements connected to each node.

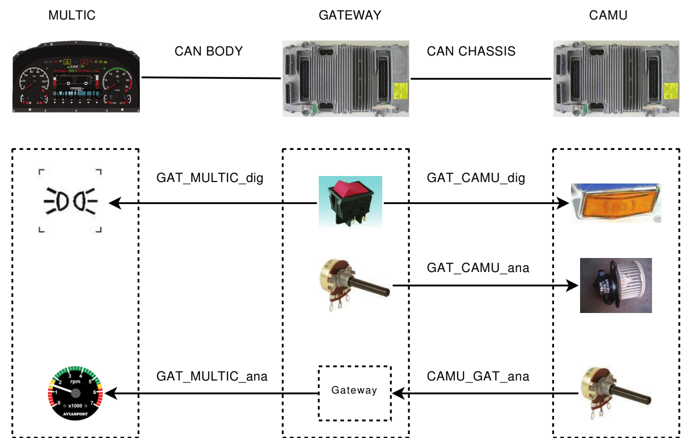


Fig. 3. Frames interchanged and node elements

Using CANdb++ software, included in CANoe tools, a database is created for each CAN network. In these databases network nodes, messages, signals and environment variables are defined. Then CANoe software is used for the simulation of the complete system. This can be done using CAPL language and using graphical panels that make possible to the user interact with network nodes. Figure 4 shows the CANoe window where the panels developed and the messages in the CAN networks can be seen.

B. *Practice about control of ECUs in a prototype*

Once designed the CAN network, the software tool can be connected to a real CAN network to check the good working of the design. The ECUs in the design are replaced for a real unit in a progressive way and the corresponding functionality checked. For this purpose a model has been designed using a dashboard and CAN units from the ACTIA manufacturer.

The figure 6 shows the implemented laboratory model for the real CAN network practice where the two CAMU units and the MULTIC dashboard are connected through two different CAN networks. The elements connected to both CAMUs are those described and simulated in the previous section: one light, one potentiometer and one fan motor in the CAMU; and one switch and one potentiometer in the CAMU gateway. The CAMUs and the MULTIC dashboard have been programmed using the manufacturer development tools named ActiGRAF and ISaGRAF whose knowledge is beyond the scope of this course. An example of a functionality implementation in ISaGRAF is shown in figure 5.

With this laboratory model and the CANoe tool the students can check and review the working of the full system and also of each part following the steps 2 and 3 mentioned in the previous section:

- Test of individual physical nodes (step 2). Each physical node is validated using a simulation of the remaining nodes. For example, for the validation of the GATEWAY node, the CAMU and the MULTIC are simulated with the CANoe while the real ones are disconnected. The inputs and outputs of the simulated nodes are accomplished

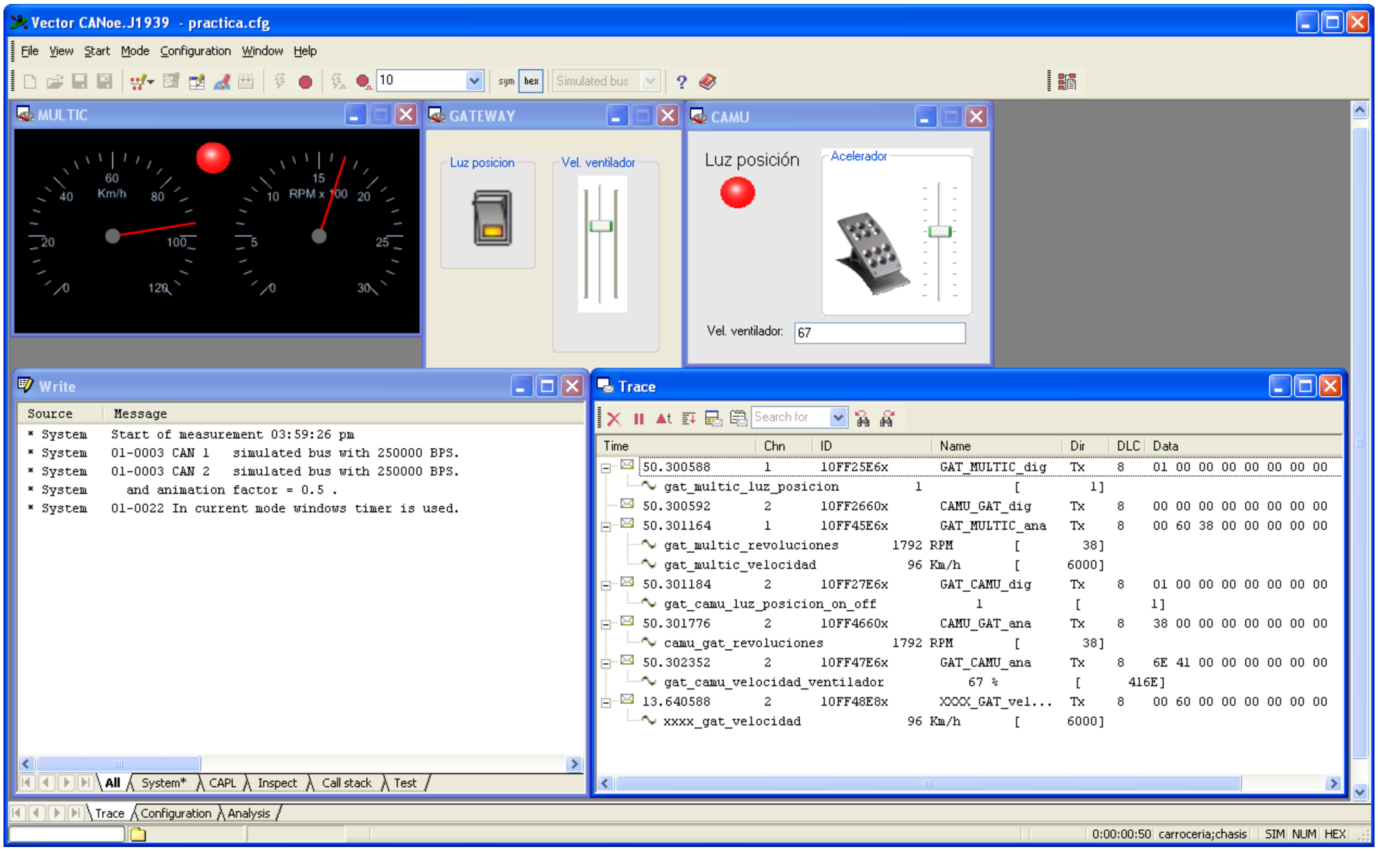


Fig. 4. CANoe main window with panels and frames

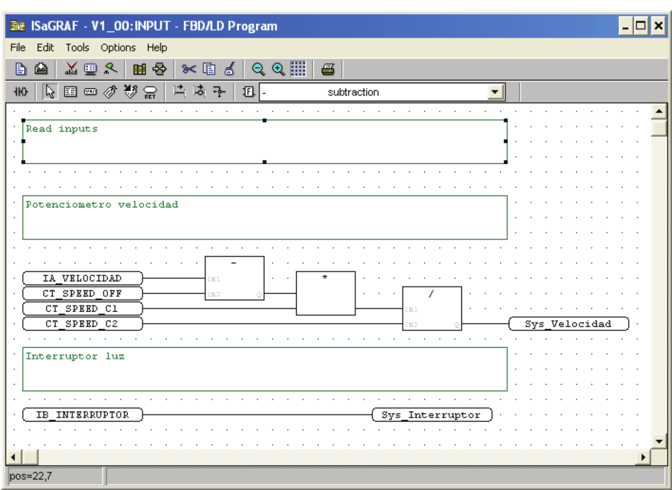


Fig. 5. Specification example of a function using the ISaGRAF tool

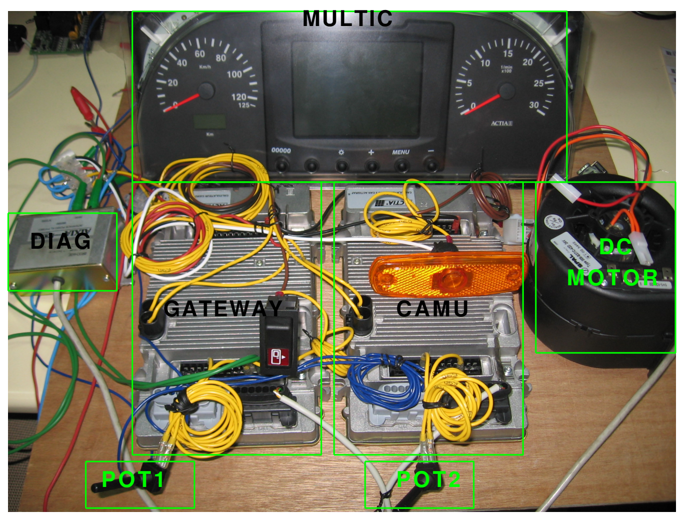


Fig. 6. Real CAN network laboratory

using the configurable panels available in the CANoe as seen in the upper part of the figure 4.

- Test of full system (step 3). Once all the physical nodes of the system were successfully tested in the previous step, the last step is to verify the right operation of the full system. This is accomplished monitoring the CAN networks (lower part of the figure 4) while operating with the real system, and at the same time comparing the data with the obtained in the simulation process.

C. Practice in a real vehicle

Finally, the students can check how the traffic (frames circulating by the communication network) of a real vehicle can be analysed using the CANoe software tool. A commercial

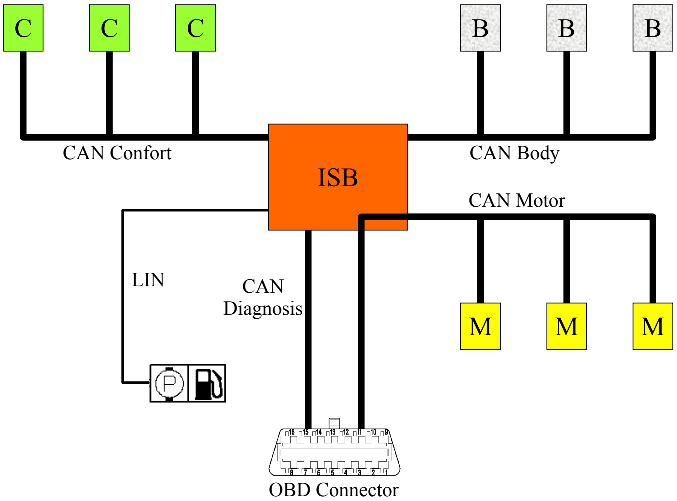


Fig. 7. Citroen C2 communication structure

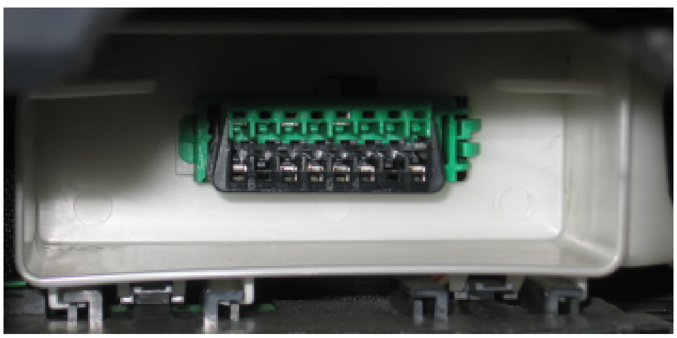


Fig. 8. OBD connector in a Citroen C2

Fig. 9. Analyzer software tool connected to the OBD connector in a Citroen C2

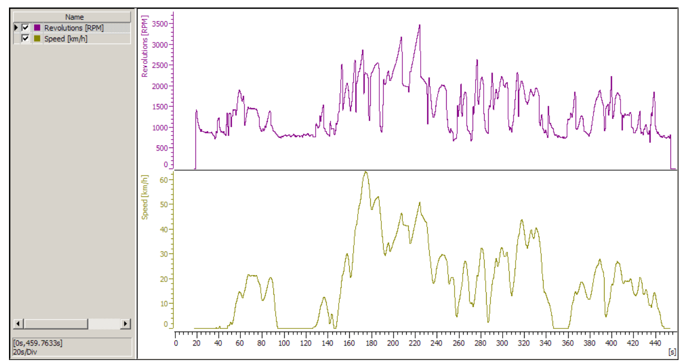


Fig. 10. Window of CANoe with the evolution of RPMs and speed

vehicle such as a Citroen C2 is used for this purpose. The structure of the communication network of this vehicle is shown in figure 7. The following networks are present in this vehicle:

- A CAN network for motor ECUs (500 Kbps)
- A CAN network for the body units (125 Kbps)
- A CAN network for the comfort functionalities (125 Kbps)
- A CAN network for diagnosis (500 Kbps)
- A LIN network for fuel additive (19,200 bps)

The ISB (Intelligent Service Box) manages the communication and the power of ECUs. The OBD (On Board Diagnosis) connector allows to connect an external tool to the vehicle, and then to communicate with all the ECUs of the automobile. It can be used for diagnosis and ECUs programming. It is a 16 pins connector.

In this practice, the students can connect the CANoe software tool to the OBD connector of the vehicle (Fig. 8) through an USB port in a laptop (Fig. 9). Thus they can see all the messages sent by the CAN networks and associate the identifiers with the signals presents in the automobile. For example, the evolution of the RPM (Revolution Per Minute) and speed can be observed such as it is shown in figure 10.

IV. Conclusion

The technologies and processes in the automotive industry are very important in the present days. Thus, it is necessary to offer an adequate specific education to obtain professionals with knowledge in these topics. The communications in the automobiles to integrate the control and management of the ECUs is essential, and the design of a set of practices about the communication networks embedded in vehicles is significant.

The authors present in this paper a laboratory test bench for training on communication networks, intended for a master in automotive technologies and processes, where the students must get a lot of knowledge in few hours. The laboratory is designed to make practices about CAN networks because it is the most important protocol used in the nowadays automotive industry.

This laboratory allows to design a CAN network and simulate the functionalities of the ECUs, using a prototype based on modules and a dashboard from the ACTIA company. Also practices in a real vehicle can be made.

The hardware and software used in this laboratory and the designed practices, allow that the students of the Master can be acquainted on communication networks in the vehicles and their role in the automotive industry, in few hours. Thus, the aim of this laboratory is achieved.

ACKNOWLEDGMENT

This work has been made in a Master sponsored by the University of Vigo, CTAG (Galician Automotive Technological Centre), CEAGA (Cluster of Automotive Enterprises in Galicia) and PSA Peugot Citroën (factory in Vigo).

REFERENCES

[1] T. Nolte, H. Hansson, and L. L. Bello, "Automotive communications-past, current and future," in *Emerging Technologies and Factory Automation, 2005. ETFA 2005. 10th IEEE Conference on*, vol. 1, Sep. 2005.

[2] J.-P. Thomesse, "Fieldbus technology in industrial automation," *Proceedings of the IEEE*, vol. 93, no. 6, pp. 1073–1101, Jun. 2005.

[3] P. Mariño, M.A. Domínguez et alii, *Industrial Communications: Basic Principles (Vol. I), Distributed Systems and Applications (Vol. II)*. Madrid: UNED, 2007.

[4] P. Mariño, *Enterprise communications: standars, networks and services*, 2nd ed. Madrid: Ed. RA-MA, 2003.

[5] K. Etschberger, *Controller Area Network: Basics, Protocols, Chips and Applications*. IXXAT Press, 2001.

[6] R. B. GmbH, *CAN specification version 2.0*. Bosch, 1991.

[7] ISO 11898, *Road Vehicles - Interchange of digital information - Controler Area Network (CAN) for high-speed communication*. ISO, 1992.

[8] ISO 11519-2, *Road Vehicles - Low-speed serial data communication - Part 2: Low-speed Controller Area Network (CAN)*. ISO, 1995.

[9] M. Estevez, "MOST and MPEG: a perfect relationship?" *Embedded Systems Europe*, pp. 36–38, October 2004.

[10] M. Prophet (editor), "Can MOST escape from the car?" *EDN Europe*, pp. 32–34, September 2004.

[11] D. Marsh (editor), "Engines of change," *EDN Europe*, pp. 58–73, July 2005.

[12] H. Arnold, "8-bit micro and LIN system basis chip combine in a single package," *EE Times Europe*, March-April 2007.

[13] B. Klosterboer, "Flexibility - A mixed signal blessing," *Electronic Design Europe*, pp. 42–44, June 2004.

[14] A. Emadi and T. Jacobius, "Interprofessional projects in advanced automotive power systems: an integrated education and research multidisciplinary approach," *Education, IEEE Transactions on*, vol. 47, no. 3, pp. 356–360, Aug. 2004.

[15] J. Dias Pereira, "A fieldbus prototype for educational purposes," *Instrumentation & Measurement Magazine, IEEE*, vol. 7, no. 1, pp. 24–31, March 2004.

[16] P. Mariño, F. Poza, M. Dominguez, and S. Otero, "Electronics in automotive engineering: A top–down approach for implementing industrial fieldbus technologies in city buses and coaches," *Industrial Electronics, IEEE Transactions on*, vol. 56, no. 2, pp. 589–600, Feb. 2009.

[17] A. Sangiovanni-Vincentelli and M. Di Natale, "Embedded system design for automotive applications," *Computer*, vol. 40, no. 10, pp. 42–51, Oct. 2007.

[18] *CANoe user manual*, Vector Informatic GmbH, 2006.

COMMUNICATIONS ANCIENT AND MODERN: HOW DOES MODERN ICT FIT INTO THE TRADITION OF PILGRIMAGES, AN ANCIENT FORM OF COMMUNICATION

Author:Peter Leonhardt (BA / MSc), Contributor: Alice Pyke (BTh)

Peter Leonhardt:BT, 81 Newgate Street, London, EC1A7AJ: peter.leonhardt@bt.com

Keywords: Pilgrimage, Spiritual ICT, Virtual Pilgrimage, On-Line Religion

Abstract: With ICT “making the world smaller”, and the need for physical travel becoming ever lower as telecommunications systems develop and spread, what are the implications for pilgrimages, an ancient form of communication? This paper examines this question, and shows how modern ICT provides a new avenue of communication to a site of pilgrimage. An avenue that can deliver services both to support physical pilgrimages to the site, and also provide services that can connect people remotely to the site in a spiritual sense. This paper examines the ICT strategies and services that are developing within this religious space and their implications, and explores how ICT can be best used by sites of pilgrimage to benefit their followers, their faith and themselves.

INTRODUCTION

Pilgrimages evolved from the desire of members of a religious faith to journey to a significant location associated with their beliefs. A pilgrimage being a communication both spiritually and physically with a geographic location. Over the last few decades, ICT has revolutionised global communication, both in the way people communicate and the speed and distance they can communicate. A general term that is used to reflect this massive impact of ICT is “making the world smaller”. ICT provides instant contact between people and places, reducing the need to travel, and the ability for anyone connected to the Internet to access and participate in global social networks for fun, religion, commerce and countless other interest groups. It seems inevitable that modern communication through ICT will have an impact on pilgrimage, this ancient form of communication.

This paper examines the impact of ICT on pilgrimages, and asks the question what are the implications of this impact both on pilgrims and sites of pilgrimage? Could ICT be a substitute for a real physical pilgrimage, thus having a direct impact on the number of people visiting a site? Does ICT offer a new channel of communication that sites of pilgrimages can offer services and support to people, providing new opportunities for people to forge links with the site? If pilgrimage is viewed as a market, how could the demand of pilgrims and people interested in a site of pilgrimage be satisfied by services offered via ICT, and who can and will try and fill this demand?

This paper attempts to answer these questions by looking at the types and methods of pilgrimage, the motivations or demands of the pilgrim and the way in which ICT can meet these demands. It also provides recommendations to sites of pilgrimage as to how to address the impact of ICT and also harness ICT as a means of support and a mechanism for extending links to people across the world.

Methods of pilgrimage

In order to understand how ICT impacts or enhances pilgrimages, it is necessary to understand the types and methods of pilgrimage that exist. Pilgrimages exist in all the major religions, with some faiths having more emphasis on the importance of pilgrimage than others. Muslims for example have an obligation to make a Hajj, a pilgrimage to Mecca once in their lifetime. Performing a Hajj is an example of the most common definition, or method of pilgrimage: “a journey to a shrine or other holy place or place venerated for its associations” (Chambers Dictionary, 1992). This is a physical connection with the pilgrimage site, with the pilgrim physically journeying to the site. ICT could clearly play a role in supporting this physical journey, providing websites with travel information and information on places to stay for example. However, pilgrimage is not simply a visit, such as a sightseeing trip, (although this may be a motivation), the pilgrim also makes a spiritual connection to the site, which could be forged through prayer or ritual at the site, or simply through thought and meditation.

A spiritual connection could be made in conjunction with a physical pilgrimage, but a spiritual connection could also be made to a pilgrimage site without a physical journey. This is known as a spiritual pilgrimage. This alternative method of pilgrimage came to prominence in the Middle Ages when it was much more difficult to make a distant physical pilgrimage to a holy site such as Jerusalem. Expense, tenure to land, difficulty and danger of travel as well as monastic vows meant a physical pilgrimage to a distant location was only undertaken by a few people (Connelley, 1999). Resources were developed to help with these spiritual pilgrimages such as interactive maps drawn by the English monk Matthew Paris in the 13th Century. These included foldouts within the map providing explanation of the spiritual significance of the locations illustrated. The Latin term *peregrinatio in stabilitate* was coined by Dom Jean LeClercq for these “meditative practices that allowed monks to make a pilgrimage with their hearts and not their feet”, (Connelly, 1999)

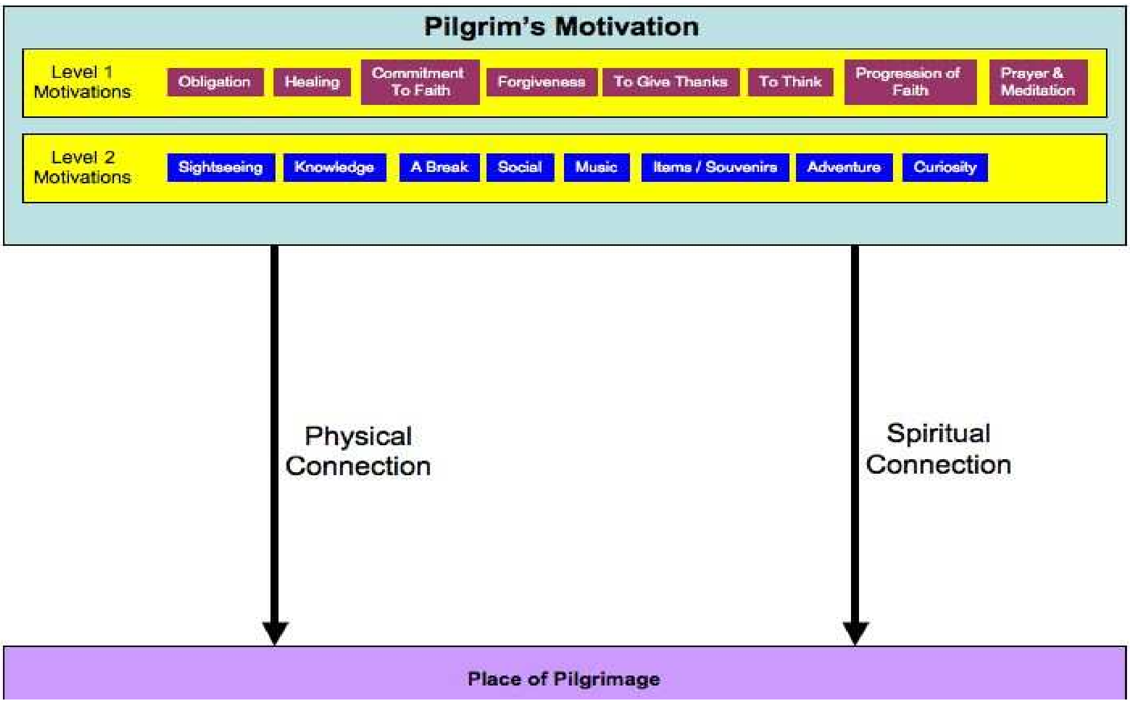


Figure 10: Pilgrim's Model showing motivational forces behind making a Pilgrimage

This distinction between physical connections and spiritual connections with sites of pilgrimage is an important point to help understand the role ICT could play in this area. Just as maps provided a point of meditation, and an interactive experience for spiritual pilgrimages in the Middle Ages, couldn't an interactive website devoted to a site of pilgrimage provide the same, or an enhanced point of meditation in the 21st Century?

In order to understand further how ICT could play a role in pilgrimage it is necessary to understand the drivers behind pilgrimage, the actual demands of the pilgrims. The next section identifies these demands in order to examine how ICT can help meet these demands.

A Pilgrimage Model

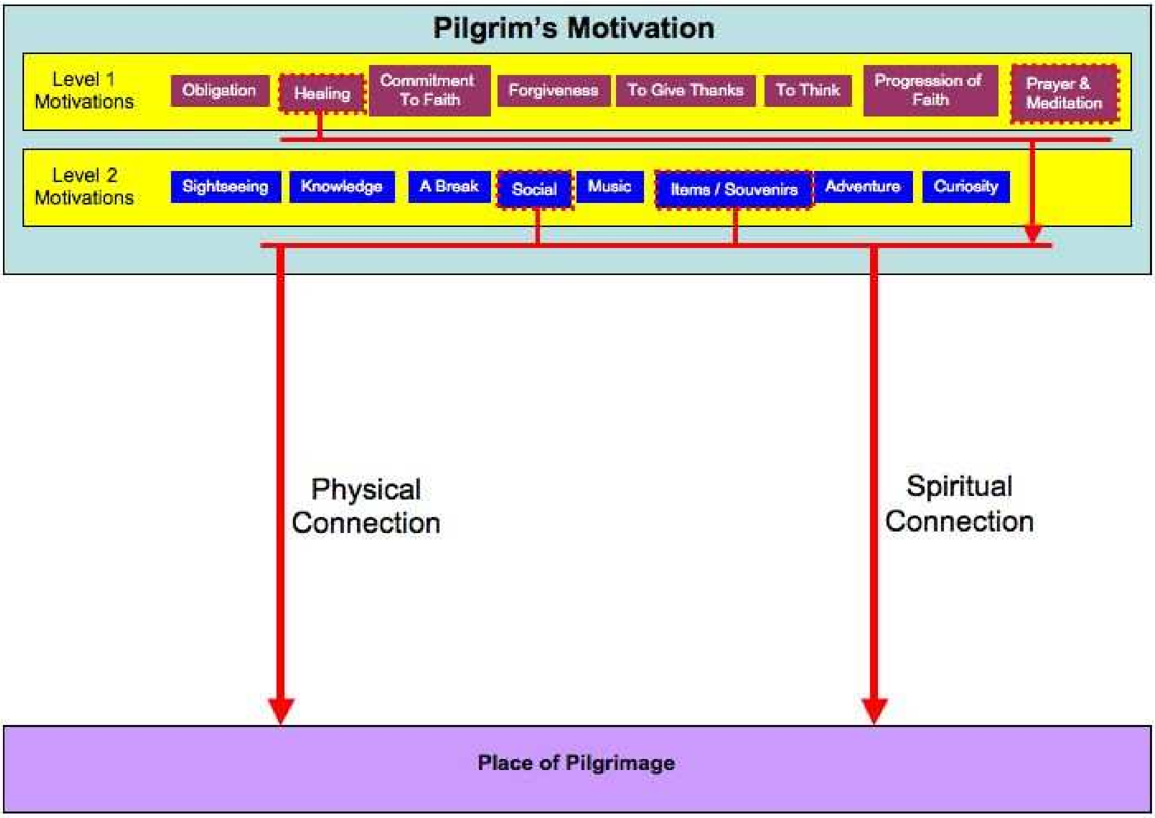
Figure 1 identifies the main motivations which drive people to undertake a pilgrimage, or form a connection with the site of pilgrimage both on a physical level and on a spiritual level. The model breaks the motivations of the pilgrim down into two levels. Level 1 are deeper, more religious or personal motivations, and level 2 motivations are the lighter and more common reasons for wanting to visit or learn more about a location. The reason for a person undertaking a pilgrimage will often be made up of more than one motivation, and also motivations from both levels.

An Example Pilgrim's Profile

Figure 2 provides a hypothetical example of a pilgrim's motivational profile for visiting the site of Lourdes in France. In this example, the pilgrim is suffering from a health problem which they hope will be relieved by bathing in the Holy water in Lourdes, and also by praying and meditating at the site. These are their level one motivations. They also enjoy the social aspect of pilgrimage, meeting likeminded people, and also would like to bring some Holy water home with them from Lourdes for anointing themselves when at home. These are their level two motivations.

These motivations, or demands of the pilgrim are met by both the physical and spiritual links to the pilgrimage site. The physical link taking the pilgrim physically to and from the site, and the spiritual link being established during the course of bathing, prayer and meditation when they are there.

Figure 11: An example Pilgrim's motivation mapped onto the Pilgrim's Model



Identifying the motivations, reasons and demands of pilgrims and general people that connect with a place of pilgrimage allows us to asses how ICT plays a part in currently and potentially assisting meet these demands. The next section examines how sites of pilgrimage are utilising ICT to do just this.

use of ict in conjuntion with sites of pilgrimage

Table 1 analyses the content and services that are provided by websites which serve some of the worlds major sites of pilgrimage, covering four of the major world religions: Christianity, Judaism, Islam and Hinduism. In most of the examples listed in the table, the website is the official website, affiliated with the pilgrimage site, the exception being The Wailing Wall, in which case the Jewish resource website www.aish.com has been used.

A pattern emerges when analysing the content and services hosted across the respective pilgrimage sites websites. All sites have some standard areas of content, in the same line as any significant place of interest. This standard content being a history of the site, with photos of the location, and the reasons why it is a sacred place of pilgrimage. Pilgrimage sites in their nature are similar to tourist locations, so the websites also generally offer travel information, when to visit, what particular events are on, and can help guide on where to stay.

So far, these areas of content are not surprising, and do not offer anything above or beyond some background info, and help with how to conduct your pilgrimage. They do not offer much beyond fulfilling an initial or mild curiosity in the site and offer help for conducting a physical pilgrimage to the site. However, what is apparent is that some websites are also offering services that are crossing into the spiritual side of pilgrimage.

	Santiago de Compostela (www.catedraldesantiago.es) **(Christian)**	**Badrinath (Hindu)** (www.badarikedar.org)	**Taize (Christian)** (www.taize.fr)	**Wailing Wall (Jewish)** (www.aish.com)	**Mecca (Islam)** (www.hajinformation.com)	**Lourdes (Christian)** (www.lourdes-france.org)	**The Vatican (Christian)** (www.vatican.va)	**Walsingham (Christian)** (www.walsingham.org.uk)
History	√	√	√	√	√	√	√	√
Photos	√	√	√	√	√	√	√	√
Webcam / TV				√		√	√	
Music	√	√	√	√	√	√	√	√
Music Scores			√			√	√	√
Donations	√	√		√		√	√	√
Souvenirs		√		√			√	
Religious Texts	√	√	√	√		√	√	√
Virtual Tour	√			√	√		√	
Prayers	√	√	√	√		√	√	√
Remote Prayers	√	√		√		√		
Social Network						√		
Email News			√					
Timetable at Site	√	√	√		√	√	√	√
Calendar of Events	√	√	√	√	√	√	√	√
Travel Info	√	√	√	√	√	√		√
Where to stay info	√	√	√	√	√	√		√

Table 1: Profiles of Sites of Pilgrimages' Websites

Spiritual ICT

Audio / Visual

Popular formats of radio and TV broadcasts are programmes from religious events and services. These can create a spiritual link with the viewer or listener, as they can sing and pray along with the service being broadcast. The Internet offers a low cost independent way of providing this live link with a site of pilgrimage. The most simple method being a webcam, such as “the wall cam” on aish.com’s website of the Wailing Wall. However, other sites such as Lourdes and The Vatican are starting to provide more TV quality content over the internet, both live and archived of significant services. Another important point to consider is that the internet provides an interactive quality you don’t get with straightforward broadcasts, and this quality allows a much more personal range of spiritual ICT services to be developed.

The Offering of Prayers, Requests and Messages Remotely

Santiago de Compostella’s, Badrinath’s, The Wailing Wall’s, and Lourdes’ websites all offer in some form, the ability for people to have their personal prayers, requests and messages offered at the site of pilgrimage via their internet site. The way this is done depends on the religion and method of prayer associated with the site of pilgrimage, and the mechanism by which ICT delivers them.

The sections below demonstrate how their mechanisms work:

Santiago de Compostela

Santiago de Compostela has two methods of making spiritual connections via its website as shown in Figure 3. These make up part of its plan to create a digital income stream that will help generate revenue to sustain the historic buildings (The Telegraph Newspaper, 2010). One method is to request a personal mass for your own reason via their internet site, which costs (at the time of writing) 7 Euros. The other method is a sophisticated system hosted by another company called mileva.com. Through this system, the user can pay 1.42 Euros to have an electronic candle lit for a finite amount of time on a candelabra displayed on a LCD screen. The idea of course is that people would like to have a candle lit within the cathedral for a personal reason, and this internet driven remote electronic candle allows a substitute for physically being there in person.

Badrinath

Badrinath, one of the key four Indian Hindu pilgrimage sites which make up the Chardham also offers Internet users the ability to make remote Pooja. Another website which offers this service is www.eprarthana.com as shown in Figure 4. Making a Pooja for a specific purpose is one of the main motivations for Hindus to make a pilgrimage to a temple. Hindu temples around India specialise in different types of Pooja, and eprarthana.com allows people to remotely perform this after paying upwards of $9(US) to make a Pooja to the temple of their choice. For example, a

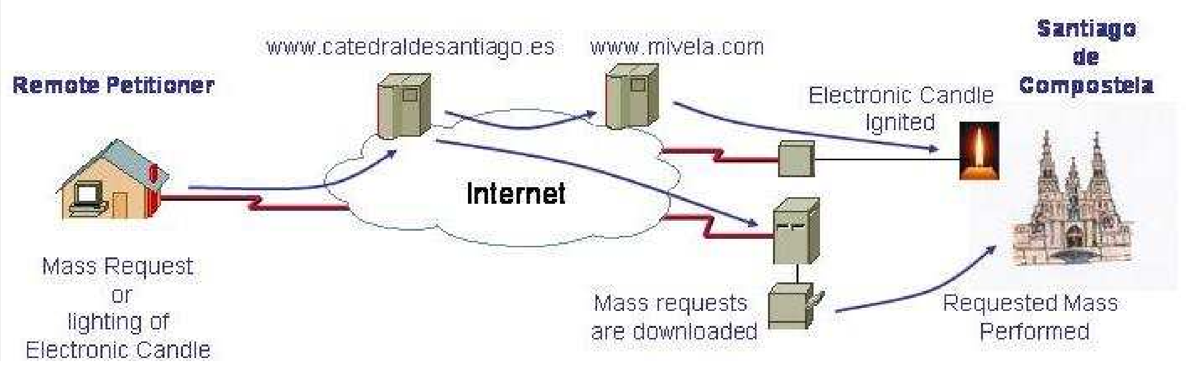


Figure 12: Spiritual Links to Santiago de Compostela

Pooja to help with debts would be made to Sri Machakara Balamurugan Temple near Chenni, or to help with health, would be made to Vaithiswaran Koli Temple. The system works by the website controlling a network of agents, which physically go and make the Pooja on behalf of the remote user.

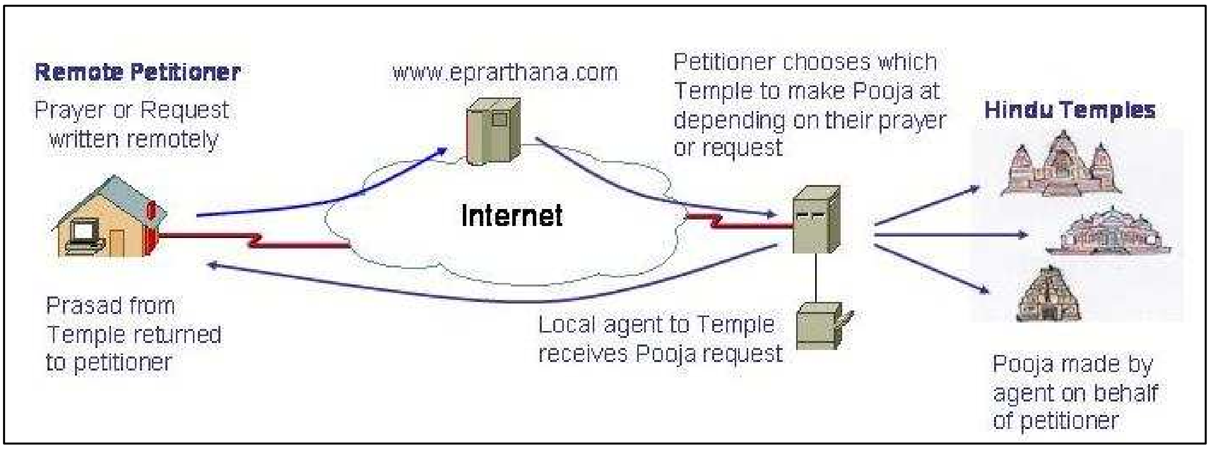


Figure 13: Spiritual Link to Hindu Temples

Lourdes

Lourdes’s website provides a service to allow internet users to remotely place a petition in the Grotto at Lourdes as shown in Figure 5. The petition is placed free via their website, then downloaded by Lourdes’ multimedia service, burnt onto a CD along with other peoples petitions. The CD is then deposited within the Grotto. This method makes an interesting case, as the petition ends up on CD, not the traditional form of spiritual communication such as text on paper or orally spoken. The digital method of storing and transmitting data is clearly seen as not a barrier to the end message finally being offered in the same format to God.

The Western Wall (Wailing Wall)

The Jewish website www.aish.com offers a service to place a note in the Western Wall free of charge, with the service being called “Window on the Wall” as shown in Figure 6. This also provides a live webcam view of the Western Wall. The internet users send a note via the website, which could be a prayer or request. The note is then printed out in the old city of Jerusalem and placed in the wall by a student of Aish HaTorah. The Window on the Wall also offers a chargeable service which allows people to place a message on the website with a backdrop of the Western Wall.

So it can be seen that ICT, especially though the Internet, is providing a new link to sites of pilgrimage, not only allowing them to provide support for physical pilgrimages, but also starting to provide spiritual connections between people and the sites also.

It is interesting to note how the websites which are offering spiritual connections to pilgrimage sites are not always officially the website of the place of pilgrimage. It indicates that there is a market for these spiritual services, and someone will satisfy the demand, and also innovate new ways to

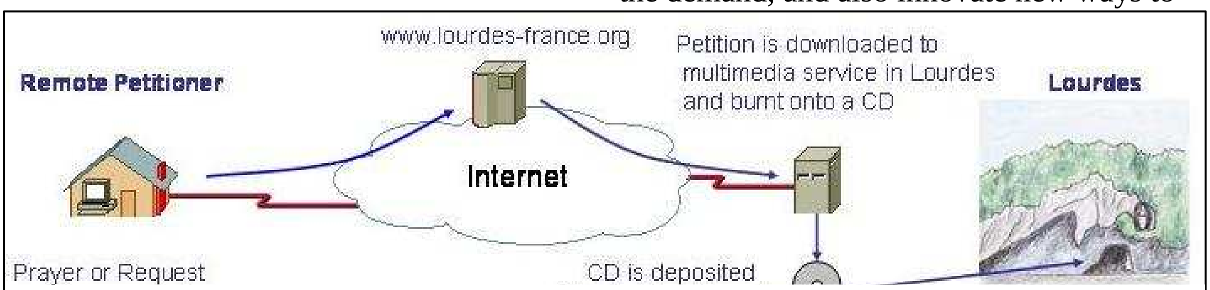


Figure 14: Spiritual Links to Lourdes

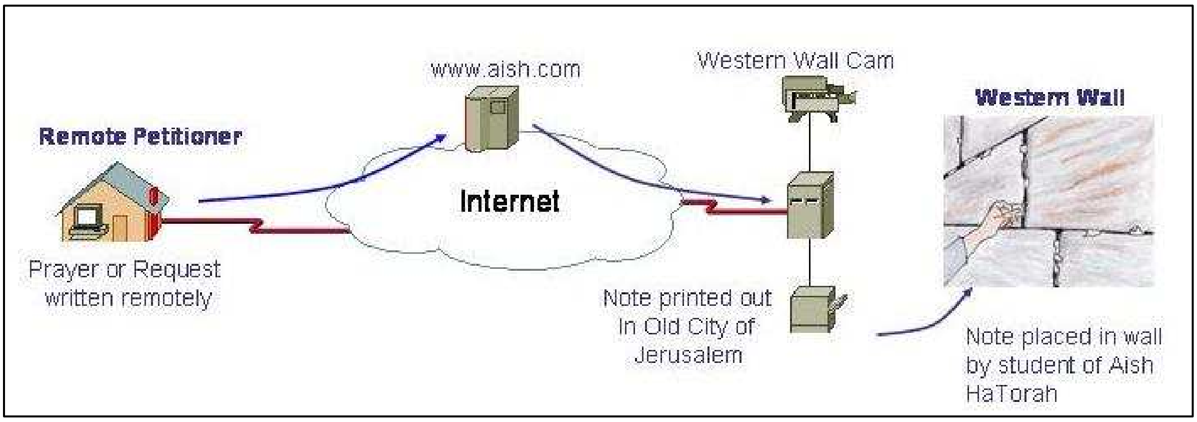


Figure 15: Spiritual Links to the Western Wall

provide these ICT based spiritual connections. It is in the interest of the site of pilgrimage to organise their own spiritual based ICT services, or refer to a trusted partner such as in the case of Santiago de Compostela referring people to mileva.com. If not, any unofficial website could set themselves up as offering such services, taking their slice of the market, and who is going to check if that remote Pooja made by someone in England was actually placed in that temple in India? There is no regulation apart from the trust and faith the real pilgrimage sites maintain, this trust and faith which should be reflected in their own websites and their own spiritual ICT services.

Social Networking

Another ICT service which is playing a part in pilgrimage and pilgrimage sites are social networking websites. There is a social motivation which plays a part in some pilgrim's drive to go on a pilgrimage, and also once people have been on a pilgrimage, they build a social network of people they have met, and a wider social network of people who have also been on a pilgrimage to the same location. Websites such as Facebook allow these social groups to be maintained in a much more active, real, and fast paced way than before. Also, they can play a role in the act of physical pilgrimage itself, such as the case with a vicar from Scotland who conducted a solo pilgrimage to Walsingham, and found that support from his followers on Facebook and Twitter via his mobile phone greatly improved his morale (Christianaid.org.uk, 2010).

Pilgrimage Model: Areas ICT Influences

In this section the Pilgrim's Model introduced in section 3 is re-examined in Figure 7 with the areas of influence by ICT mapped onto it as identified in the preceding section.

The use of ICT through the Internet can help satisfy the motivations of pilgrims, both directly and indirectly, and influences both the physical and spiritual links to the site of pilgrimage. In terms of ICT roles on the physical link, it plays a support role, an indirect role. ICT can provide information on the site of pilgrimage in terms of how to get there, where to stay, and what's on when. ICT supports pilgrims planning on making a physical journey to the site.

A website for the site of pilgrimage can also help satisfy to a certain level some of the level 2 motivations shown in the model. Sightseeing, knowledge and curiosity can all be catered for through the use of photo galleries, text and virtual tours. Virtual tours are becoming increasingly sophisticated, utilising content rich web based media players and exploiting the increased bandwidth typically now available to the user. Just how far a virtual tour could go to replacing a real tour is a matter of debate and personal opinion, but a virtual tour can certainly act as a showcase for the real thing, encouraging people to interact with the site further. In some cases a virtual tour may be the only option if they can't travel through lack of money or health.

Other level 2 motivations which can be catered for, or enhanced to a certain degree include social networking (as discussed in section 4.2.5) and obtaining items from the site of pilgrimage through simple e-commerce mail order. Music associated from the pilgrimage site is also readily available on some of the sampled sites, mostly on MP3 form available to download, but also in some cases in full multipart music scores, along with tutorials on how to sing or play the pieces. This is the case with Taize's website, which provides the full music score to all of their Taize chants. The Taize chants have become one of the keys things associated with the site in France, and one of the reasons people visit. By providing the ability for musicians to learn and play their chants over the Internet, Taize can also spread their methods of worship and in turn, raise peoples awareness of Taize and raise awareness of their cause which is to build relationships between different Christian groups to avoid future conflicts.

Level 1 motivations can also be catered to a degree through ICT, again, with the Internet providing the interactive link between the site of pilgrimage and the pilgrim. Prayers and offerings can be made, which may satisfy some peoples motivations to visit a site, or perhaps just help maintain the spiritual link to a site of pilgrimage which they have visited in the past or intend to visit in the future. The interactive quality of the Internet, and the increases in bandwidth and media services supported by browsers and web servers also now allow more audio / visual rich content to be streamed to the user than ever before. All this may help build the spiritual link between the person and pilgrimage site.

Through the use of ICT, and web based services, a site of pilgrimage has the possibility of providing a wide range of services which can both assist with both physical pilgrimages and create and maintain spiritual links, as well as publicise themselves and their causes. It also offers the site of pilgrimage the option of a new digital revenue stream.

Where Could the future lead?

The obvious question to help answer where the future will lead is to look at the number of pilgrims visiting sites of pilgrimage now compared to the past. Examining the statistics for Santiago de Compostela between June 2004 and June 2010 makes a good case study, as both of these years were “Holy Years”, when June the 25th falls on

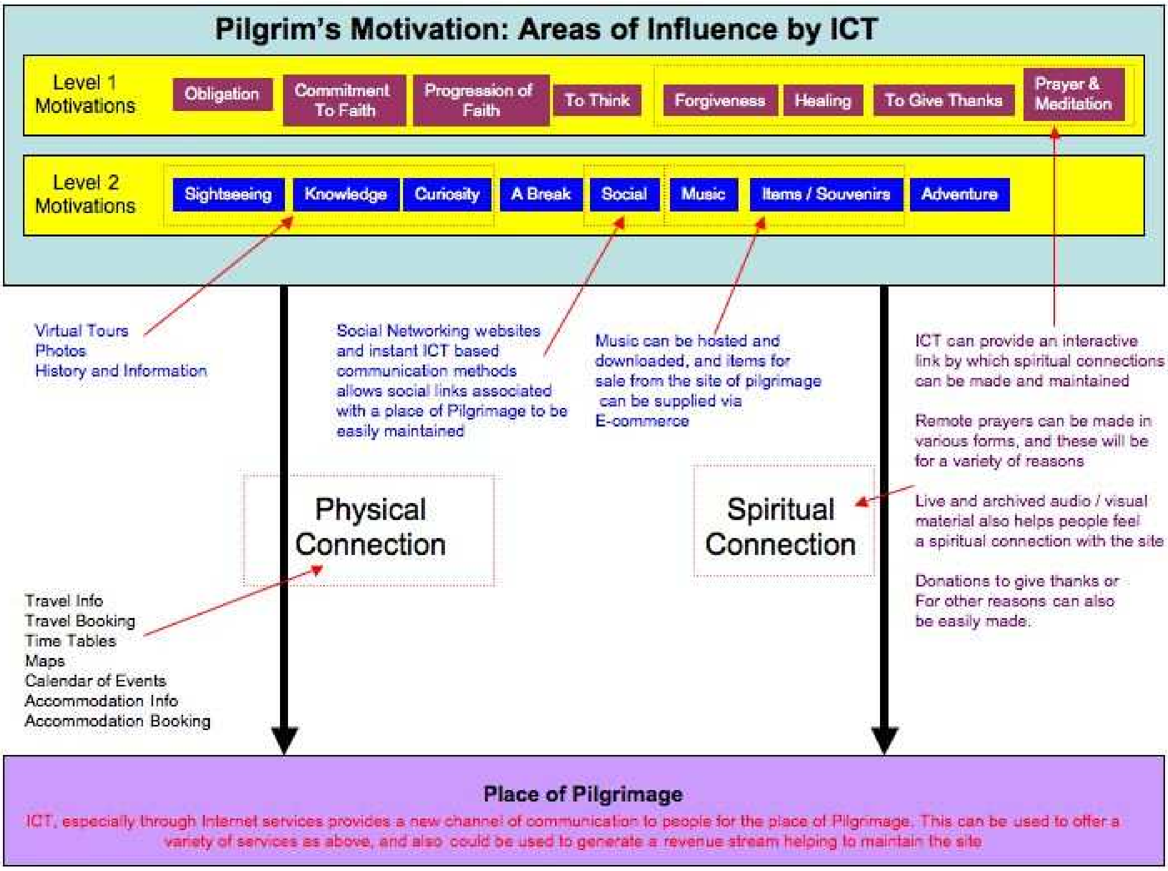


Figure 16: Pilgrims Model showing Areas of Influence from ICT

Sunday and visiting the site becomes more attractive. Also, by June 2010, Santiago de Compostela's web site and spiritual ICT services were in full swing compared to 2004. In June 2004 19,924 pilgrims were received at the pilgrims office, in June 2010 33,721 pilgrims were received (Peregrinossantiago.es, 2010) That's an increase of 69%! Clearly there is no indication from these figures that a virtual tour, or a remote prayer is a substitution for a real physical pilgrimage in this case.

What this does show is that not only is the interest and participation of making pilgrimage to Santiago de Compostela being sustained, but that it is actually growing. This interest and demand in pilgrimages is part of a very large general market for religious books, music, holistic products and religious travel estimated to be approx $40 (US) billion (Godweb.org, 2010). With this huge demand for spiritual goods and services it is inevitable that we will see new and innovative ways to supply the demand using ICT and the Internet as we have seen ICT and Internet cement its role in almost all industries over the last 15 years.

There will be two broad types of services which pilgrimage sites will be able to offer: physical support services for making a real pilgrimage and spiritual ICT services. Physical support has already been developed, in much the same way as the tourist industry, but spiritual ICT has a long way to go.

As bandwidth increases, voice, video and data converge over IP, and peripheral equipment develops the ability to offer new IP based services, there is no doubt these advances in technology will be utilised for more content rich, interactive spiritual ICT services. Combined with the Internet generation growing up, and utilising other types of web services such as Facebook for forming, maintaining and organising groups, all these factors will play a role in how people will interact with sites of pilgrimage.

Can a virtual pilgrimage ever replace a real pilgrimage? There is no evidence to suggest this at the moment. A real pilgrimage also involves a journey, not just a visit, and being able to simulate a journey, with all its events, meetings, ups and downs and physical activity is not something a virtual pilgrimage could reproduce. Being part of a group both on the journey, and also at the destination with all its spiritual significance, along with the communal worship and rituals, singing and social interaction are all important factors which make up a pilgrimage, and not something which can be replaced though ICT services…at least not yet!

Maybe the next step for spiritual ICT is to set up servers within the walls of sacred sites, a consecrated data centre. In much the same way as sacred texts store information such as the Bible, a server can also be seen as simply a store of information, so, in principle there would be opportunities here to offer spiritual ICT services with even stronger roots to the sacred site it is located in.

Spiritual ICTs' role at the moment is an extra service to the actual site of pilgrimage. People already with an affinity with the pilgrimage site are most likely to use the spiritual ICT services it offers, to maintain links, provide support, provide a focus of meditation, and also provide a place of stability and refuge in an every faster paced and changing world. Here lies the strength of real, physical, established sacred sites of pilgrimage. In the digital age of virtual meeting places, online shops and homeworking, people want a real place they can connect with, not just a server in cyberspace.

conclusions

1 – There is a role for ICT to play in pilgrimages. Its role can be divided between providing assistance to real physical pilgrimages, and providing services which are spiritual in nature, connecting people virtually to a site of pilgrimage. A further role of ICT is in its ability to create and maintain social networks, both through simple group emails or more sophisticated social networking websites. This allows the social links between people forged via pilgrimage to remain active.

2 - It is possible through the Internet, and other peripheral applications such as mobile phone apps, to provide ICT services that are spiritual in nature. The interactive capability of ICT means it is capable of delivering a more personal connection between people and a spiritual event or location than one way broadcast methods such as TV.

3 - A site of pilgrimage acts as a foundation from which spiritual based ICT services can be provided. A site of pilgrimage adds trust and credibility to spiritual ICT services, as well as fulfilling a demand of pilgrims and other people who want to create and maintain remote spiritual links with the site.

4 – The current role ICT plays in pilgrimage does not threaten the continuation of the traditional form of pilgrimage, if anything, pilgrim numbers are increasing such as in the case of Santiago de Compostela. Perhaps the pilgrimage site's Internet sites are acting as marketing, encouraging more people to visit by providing the history, religious roots and virtual tours of the often spectacular and spiritual locations.

5 - As services, types of content and bandwidth increases over the Internet, the opportunities for even more immersive interactive experiences between people and sites of pilgrimage will grow. One of the areas that will benefit from these developments in technology and telecoms infrastructure will be spiritual ICT, and spiritual ICT services will grow in proportion to physical support services offered from pilgrimage sites.

recommendations

1: To Sites of Pilgrimage: Sites of pilgrimage should consider setting up a website, and other ICT applications offering official spiritual based ICT services if they have not done so already. An official website owned by the site of pilgrimage means they have control over the types, method and payment of the official spiritual ICT services offered in their name, or even a statement on their policy regarding such services. If the site of pilgrimage does not set this up, then a vacuum in the market is left which will be filled by the private sector, which means loss of control and revenue by the pilgrimage site over their own assets.

2: To ICT Solutions Providers: There is an opportunity to partner or offer to sites of pilgrimage an ICT strategy and service based on the methods and opportunities outlined in this paper. A site of pilgrimage is unlikely to have the expertise or time in house to deliver and maintain an ICT solution that utilises the growing range of multimedia services available over the Internet, so working with an ICT solutions provider would be in their interest.

REFERENCES

Chambers Dictionary , 1992 W & R Chambers Ltd

The Telegraph Newspaper, 03 May 2010. *Santiago de Compostela welcomes armchair pilgrims*

www.Christianaid.org.uk, June 2010. *Christian Aid's Website*

Connolley, D.K., 1999. *Imagined Pilgrimage in the Itinerary maps of Matthew Paris*, College Art Association.

http://peregrinossantiago.es, June 2010. *Pilgrims Office of Santiago de Compostela*

www.catedraldesantiago.es, June 2010. *Santiago de Compostela's official website*

www.mivela.com, June 2010. *"My Candle" website*

http://www.godweb.org, June 2010.

www.eprarthana.com, June 2010, *Virtual Pooja*

www.aish.com, June 2010, *Window on the Wall*

www.lourdes-france.org, June 2010, *Lourdes Official Website*

www.badarikedar.org, June 2010, *Badrinath's official website*

Manuscript, 1273, Matthaei Parisiensis Historia Anglorum, Propria, Nanu Exarta.

www.taize.fr, June 2010. *Taize's Official Website.*

www.hajinformation.com, June 2010, *The Harj Official website (English)*

www.vatican.va, June 2010. *The Vatican's Official Website*

www.walsingham.org.uk, June 2010. *Walsingham's Official Website*

THE JUMP
What is next

Antonio Luis Flores-Galea
MS Communications & Electronic Engineering, MS International Trade
ICT Consultant – Ares 6, E9, 2-2, ES41014 Seville (Spain)
aflores@ieee.org

Keywords: web 2.0, blog, 3G, LTE, *Creative Commons*, pay-for-use, augmented reality, social habits, Internet, change

Abstract: *The jump* can be better described by its consequences. All these years we've been used to *receive* information from media like radio, TV or newspapers. Currently all the people have the opportunity to *send* information to everyone, everywhere. This fact is changing our social habits and will cause all topics involved in information processes be reconsidered. This article analyzes the reasons for *the jump* and compares how did it work before it and how will it probably work afterwards, in terms of technology and business models: the blog phenomenon, 3G and 4G streaming tools, *Creative Commons* licenses, pay-for-use model and augmented reality will be discussed among radio adaptation to the current needs and the way TV is mixing with Internet to produce a new way of entertainment and information system. And, on the other hand, a social study of behaviours and trends since broadcasting was able until now is provided as well.

1 INTRODUCTION

Every person on Earth has always felt the need for information. Information was the base of the society even before the History began. Nowadays it's probably the main concept for evolution and progress worldwide.

Broadcasting was the first tool for spreading information to any location on Earth at the speed of light and it was one of the very first steps forward the current technological world.

However, *the jump* has recently broken the line of business based upon broadcasting, the way of doing things and will probably change the whole conception of all topics related in information processes we currently deal with.

2 Send Vs. Receive

The jump can be better described by its consequences. All these years we've been used to *receive* information from media like radio, TV or newspapers. Currently all the people have the opportunity to *send* information to everyone, everywhere. Neither of those media is able to provide the user with a return channel for sending information by itself. But the Internet and the telecom carriers networks do. The web 2.0 applications and those for mobile devices are shaking the traditional broadcasting model.

3 Behaviours

As a result, *the jump* is based upon technology but is not about technology. *The* *jump* consists of a social revolution, where individuals have jump out to the scene to play like a kind of traditional journalist.

Nonetheless there is a difference: no one pays for that. This new model relays on doing things in the subject's own interests, the love for sharing the information and the concern of establishing relationships with others.

3.1 The teenagers' jump

The widest poll for teenagers' habits there currently exists, surveyed over 10,000 teens aged 13-19 who were the 15% most active social networkers on myYearbook —the #1 site for teens in the US according to comScore—, offers interesting results to be analyzed here. These are shown in the figure 1.

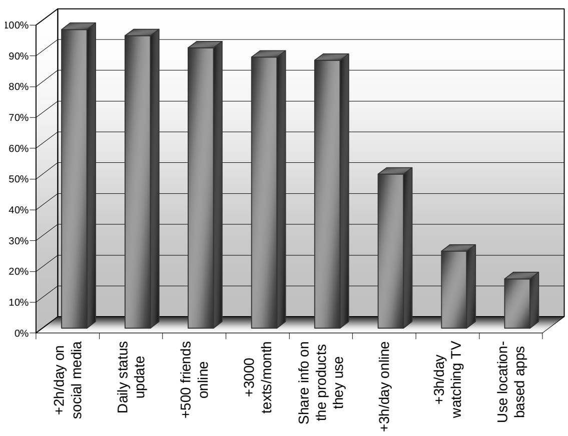


Figure 1: Teenagers' current behaviour about technology.

The first conclusion that shows up is the generalization of social media networks among them. These websites manage a huge amount of information, including texts, photos, video, links and almost every kind of content anyone would share with others.

On the other hand, location based applications stand for the loser, as they cause the individuals to feel nervous or annoyed.

The third idea we should take into account is the fast crash the traditional social media are currently undergoing. Attending on these results, it could be said two times more people are preferred to be online than watching TV.

3.2 The *silver surfers*

This is the name for people older than 50 which are active on the Internet. A recent study from the AARP —the American Association of Retired Persons— shows that over 27% of them uses the social media to stay in contact with their family. 40% of all adults aged 50 and older said they consider themselves extremely or very comfortable using the Internet. They use the social web to keep in touch with family members. Almost 75% said they are connected online to relatives other than children and grandchildren, 62% are connected to their children and 36% are connected to grandchildren, as well.

This is also part of the jump, as many of these older people will spend on high-tech gadgets, if they find them useful. The more global the world is, the more attractive the new Internet era becomes.

4 Technologies

The jump is based upon several technologies developed in the last years, although it's further a social phenomenon.

Nevertheless it's necessary to understand the basics of them to understand the jump itself.

In this chapter the most influencial technologies in the author's point of view are shown.

4.1 The Web 2.0

The first time this term was used was in the O'Reilly Media Web 2.0 conference in 2004. The term *Web 2.0* is commonly associated with web applications that facilitate interactive information sharing, interoperability, user-centred design, and collaboration on the World Wide Web.

A Web 2.0 site allows the user to interact with the page as a contributor for its content, in contrast to classical websites where he or she is limited to the passive viewing of the information.

Web 2.0 websites typically include some of the following features. Andrew McAfee used the acronym **SLATES** to refer to them (McAfee, 2006):

- **Search**: Finding information through keyword search.
- **Links**: Connects information together into a meaningful information ecosystem using the model of the Web, and provides low-barrier social tools.
- **Authoring**: The ability to create and update content leads to the collaborative work of many rather than just a few web authors. In wikis, users may extend, undo and redo each other's work. In blogs, posts and the comments of individuals build up over time.
- **Tags**: Categorization of content by users adding *tags* —short, usually one-word descriptions— to facilitate searching, without dependence on pre-made categories.
- **Extensions**: Software that makes the Web an application platform as well as a document server.
- **Signals**: The use of syndication technology such as RSS to notify users of content updates.

The most common browser technology in Web 2.0 sites is AJAX (Asynchronous JavaScript and XML), followed by Flex from Adobe. AJAX code allows data on a web page to be downloaded and uploaded from the server without undergoing a full page reload.

The data fetched by an AJAX request is typically formatted in XML or JSON (JavaScript Object Notation) format, two widely used structured data formats. Since both of these formats are natively understood by JavaScript, a programmer can easily use them to transmit structured data in their web application. When this data is received via AJAX, the JavaScript program then uses the Document Object Model (DOM) to dynamically **update** the web page, allowing for a rapid and interactive user experience. This is how the page can act like a desktop application on the browser.

One of the success keys for this technique to be applied is the possibility to have *widgets*, which are independent pieces of code that accomplish such a common task as picking a date from a calendar, displaying a data chart, or making a tabbed panel.

On the server side, the jump has been able by using standard formats to share the data saved in files and databases with other websites. In the early days of the Internet, there was little need for different websites to communicate with each other and share data. Currently sharing data between sites has become an essential capability (think of the Facebook *like button*). To share its data with other sites, a web site must be able to generate output in machine-readable formats such as XML, RSS, and JSON. Imagination for developing new services based upon Web 2.0 applications do the rest.

4.2 3G/4G Services

Some years ago 3G allowed us to access to Internet through a wireless link globally, and this fact has produced a wide set of mobile devices in the market, such as *smartphones*, *netbooks* and, more recently, *tablets* and *ereaders*.

Now 4G services are about to be launched. They're based on the *Long Term Evolution* (LTE) protocol, a new high-speed radio access method for mobile communications systems. LTE offers a smooth evolutionary path to higher speeds and lower latency, realising the full potential of LTE requires an evolution from today's hybrid packet/circuit switched networks to a simplified, all-IP (Internet Protocol) environment. From an operator's point of view, the pay-off is reduced delivery costs for rich, blended applications combining voice, video and data services plus simplified interworking with other fixed and wireless networks.

LTE is intended to be capable of supporting the broadband Internet user experience we already enjoy in today's fixed networks —with the addition of full mobility to enable exciting new service possibilities.

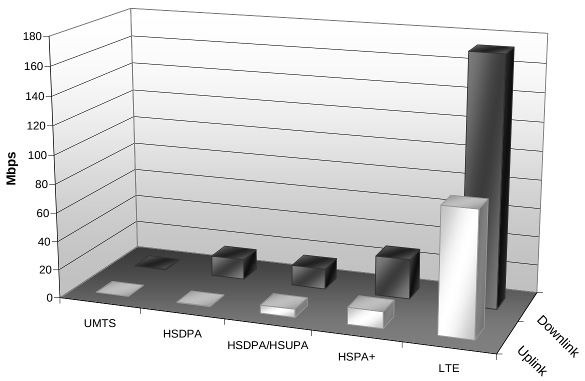


Figure 2: 3G and 4G technologies bandwidth.

Some services to be considered are large-scale streaming, downloading and sharing of video, including high-quality videoconferencing, downloading and sharing music and rich multimedia content.

All these services will need significantly greater throughput to provide adequate quality of service, particularly as users' future expectations will be increased by the growing popularity of other high-bandwidth platforms like High Definition TV transmission.

LTE will soon bring the features of Web 2.0 into the mobile environment, as well as real-time peer-to-peer applications like multiplayer gaming and file sharing.

In addition, some new applications are to be deployed, including 'machine to machine' (M2M) interaction sharing images or real-time controlling, Internet connected vehicles and large-scale exchange of information within community-based projects.

4.3 Augmented Reality

Augmented reality (AR) consists of adding extra information or images over a real image in real-time, like in Figure 3. It's been widely applied in sports matches on TV.

This technology is based upon several features:

- Live video streams available to be processed.
- Enough processing power on a device, usually a mobile one.
- A range of sensors, like a GPS receiver, a compass, gyroscopes and accelerometers.
- A database with the information to be searched and showed.
- The AR software application that generates the final view for the user.
- A way to display these results.

Figure 3: Application for augmented reality.

The most genuine applications currently available are:

- Labels to be displayed on parts of a system to clarify operating instructions for a mechanic who is performing maintenance on the system.
- Images of hidden parts, which can be particularly effective for medical diagnostics or surgery.
- Safeguarding for confidential data in combination with existing real prototypes.
- Layering of virtual maps over a terrain for military or emergency services, as well as for hydrology, ecology and geology study.
- Simulation of planned construction for architects.
- Showing ruins, buildings, previous landscapes or decoration rebuilt, when sightseeing.
- Videoconferencing with real and virtual participants.
- Guiding lines directly on the road for GPS navigation.
- Adapting views for panoramic virtual windows in homes or offices.
- Virtual gadgets, like clocks, consoles, calendars, etc. on the wall or the table.
- Feeds for augmented reality data and user information geo-posts (for example, virtual graffiti or virtual artwork analysis in a museum).

4.4 Electronic paper

Electronic paper is a new technology based on electrochemical processes. It will replace the traditional ink and paper in those applications where content or colour updating is expensive.

Currently, the most common utility are *ereaders* but this is only the beginning of *the jump*. Electronic paper will soon change our behaviour, our tools and the society itself (Heikenfeld, 2010).

Nonetheless there are high-quality, full-motion displays based on back-lighted LCD. They are quite affordable too. Then why do we need electronic paper displays?

Electronic paper does not consist of light filters like LCD but real ink, therefore it attends on reflection instead of refraction. The advantage is precisely this: a backlight source is not needed hence power consumption is dramatically lowered and it can be used under the sunlight, where an LCD backlight fails.

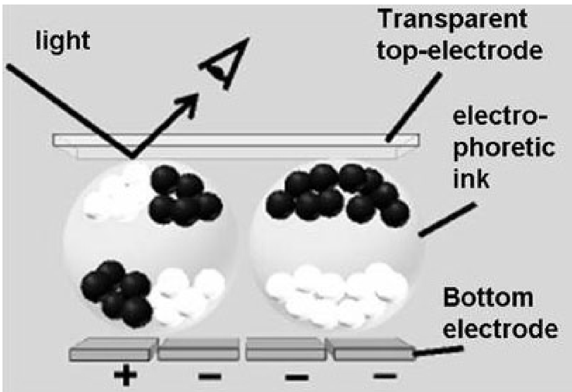


Figure 4: B&W electronic ink principle.

A very interesting application is electronic paper labels for stores, as they reduce the cost of updating prices, taxes or special deals but also let the store implement a price policy for variable prices upon the day of the week as well as the hour of the day. There could be possible hence a price policy in terms of the current sales in an specific store department or the remaining stock.

Another application to come is the use of transparent electronic paper on smart windows able to shade the light passing through them and turning them opaque for privacy. Even though there is a more interesting thing the electronic ink can deploy. It is the ability to act as an IR or UV filter so it can protect our home furniture or a museum collection, as well as reduce overheating due to direct sunlight in summer.

Last but not least, attending on its low power consumption and high visibility, another usage comes up with traffic panels. Additionally electronic ink could also make them adaptable to the traffic or the weather condition easily and inexpensively.

5 The Contents Era

Information is currently the basis for the operation of the world. Like a Swiss clock machinery, the global economy ticks with all parts of business doing it possible together and instant information is what does make it work.

The content industry recent developments have attached importance to the fact that information is a real and very valuable part of a company assets and therefore individuals should realize it is one of the most powerful tools one can manage nowadays.

It is certainly this point where the jump has push up all the people to generate information and this information has become a need for the society to work.

The rules are not the same as ten years ago. The most important progress on that topic can be found in this chapter.

5.1 *Blogs*

Blogs are probably one the most impressive changes on communications occurred in the last decade and the clearest expression of the jump.

A *blog* is nothing but a website where its author publishes articles (called '*posts*') regularly, displayed in reverse-chronological order. Currently these posts are being replaced by much shorter *tweats* (Twitter website updates) in some cases so one should think about what the future of *blogs* and content sharing will be. Two scenes can be considered: evolution and extermination (or darwinism) (Rubel, 2009).

One possibility is that *blogs* remain the primary social hub for many of the people, hosting all kind of content, aggregated from anywhere (Posterous is a good example). Another possibility for evolution is that *blogs* become social networks of their own (Facebook is getting linked with more and more *blogs* every day).

On the other hand, there's no doubt that the micro-*blogs* and social networks are stealing time away from *blogs*. Given our finite window of attention this may threaten many existing *bloggers*.

5.2 *Creative Commons* Licenses

Creative Commons (CC) licenses were created to allow the authors other ways for copyright management besides the classical "*All rights reserved*" statement. Therefore these licenses stand for "*Some rights reserved*". They can be understood as the GNU license for content. As an example, anyone interested in distributing his intellectual work can use them to force the author to be mentioned despite the work is available for free. Original licenses are:

- **Attribution (*by*):** Licensees may copy, distribute, display and perform the work and make derivative works based on it only if they give the author or licensor the credits in the manner specified by these.

- **Non-commercial (*nc*):** Licensees may copy, distribute, display, and perform the work and make derivative works based on it only for non-commercial purposes.
- **No Derivative Works (*nd*):** Licensees may copy, distribute, display and perform only verbatim copies of the work, not derivative works based on it.
- **Share-Alike (*sa*):** Licensees may distribute derivative works only under a license identical to the license that governs the original work.

The high volume of information uploaded to the Internet by users in the last few years has encouraged the intensive use of CC licenses. Many people generate content but very few of them want to get cash for it, so CC is the ideal solution.

But what moves people to share information, to develop full free sites or applications, to upload tons of photos and video with no money compensation? The answer is found in psychology, next chapter.

5.3 Tiny & Specific: *Widgets*

Everyone wants to be known for anything. The Web 2.0 offers the possibility to create your own fan group despite it is very small. This fact makes a person to feel loved and important and, at the end, this is the source for the Web 2.0 explosion.

The more impact you want for your idea the more interactive and easier to share your 2.0 application should be on the web. Some interesting evolution for easy sharing of applications in the Web 2.0 is the concept of '*widget*'. A widget is a piece of interactive code that can be inserted into an webpage or even the computer desktop easily —usually just drag and drop— to perform a very specific task, like a clock or a street map.

5.4 Pay-per-use model and monthly fee

Nowadays, due to the impact of new generation global communication systems, specially the Internet, a lot of business are to be redefined. That's a part of the *Business Model Innovation* (BMI) concept (Christensen, C., 2008).

Probably the first concern is about determining a pricing strategy. Two primary categories of pricing taxonomy are *pay-per-use* (each time one uses a service a well known amount is charged). and *fixed recurring pricing* (the user pays some periodic fee for generally unlimited access). These two strategies take there isn't a product but a service for granted. This mean making a tough decision for some traditional business and this is indeed part of the jump.

The decision about selecting one of them relays on the value acquisition from the customers perspective. In order to charge on a recurring basis, the customer has to feel the need for using it frequently. If the service is a very valuable solution to be used once or twice a month, you'll barely find anybody interested in a monthly fee. Most of the companies actually offer both models for their services, hence the user is able to switch between them as needed.

6 CONCLUSIONS

All we have read in this article is about the changes caused by the last technology available in the market nowadays. It also shows the path for the next evolution in personal and business communications. The jump, as described, is mainly based on the Web 2.0 concept and therefore on the interaction between the subjects and the Internet. Spontaneous and scattered information generation is the base of our current world and the principle for the progress in the future. And, attending on the increasing change in our behaviour and environment, you can figure almost everything is to be done yet.

ACKNOWLEDGEMENTS

Enrique Dans. *www.enriquedans.com*.
Ana Maria Trigo. *www.anatrigo.es*.
IEEE TMC Spanish Chapter.

REFERENCES

McAfee, A., 2006. *Enterprise 2.0: The Dawn of Emergent Collaboration*. In *MIT Sloan Management review*. Vol. 47, No. 3, p. 21–28.

UMTS Forum, 2008. White Paper on *Towards Global Mobile Broadband Standardising the future of mobile communications with LTE*.

Heikenfeld, J., 2010. *The Electronic Display of the Future*. In *IEEE Spectrum* (Vol. 47, March 2010).

Rubel, S., 2009. *Mashable Mind Map: What is the Future of Blogging?* In *www.mashable.com*.

Christensen, C., 2008. *The Innovator's Prescription*. McGraw Hill.

SMARTER BUSINESS MODELS FOR SMART LIVING, SMART COMPANIES AND SMARTER SOCIETIES

Mauro Ugolini, Patrizia Macrina, Jonathan Buschman
Ericsson Telecomunicazoni Italia S.p.A., Via Anagnina 203, Rome, Italy
mauro.ugolini@ericsson.com, patrizia.macrina@ericsson.com, jonathan.buschmann@ericsson.com

Keywords: Smart Cities, Business Models, ICT revolution, Sustainability, Coopetition, Smart Metrics

Abstract: Even as traditional business models are challenged more than ever by diverse threats, unlocking the full potential of the broadband revolution offers a real opportunity to take companies' business to new levels. What are needed are new, innovative business models that will make possible unleashing the strengths of the information technology revolution, enabling greater sustainability and real growth. We show that programs enabling the evolution towards a smarter model of society and based upon cooperation among universities, ICT industry, public administrations, and SMEs require a change in the way of doing business. Companies that do not understand this and do not adapt will not be positioned to take advantage of new, coming, business opportunities, driven by the information technology revolution. The main features of our work are a focus on increasing the economic prosperity of a region, trialling of new business models, development of innovation ecosystems involving heterogeneous players (industry associations, private/public financers, and local public administrations), and measurable results in terms of impact on regional growth and improvement in the quality of life. Our model involves a staged approach. We describe in detail this model.

1 INTRODUCTION

According to the early 20th century economist Joseph Schumpeter both growth and crises in the capitalist system are the result of technological innovation (Schumpeter, 1911) – more precisely, the upswings and downswings in growth (Schumpeter's "long waves") are caused by the clustering of innovations during technological revolutions.

A worldwide proponent of the neo-Schumpeterian theories, Carlota Perez, has recently reviewed Schumpeter's work (Perez, 2002), and focused the investigation on the economic/social impacts of the new technologies, finding that at every technological revolution there is "a shift of business climate" – a "great surge of development", as she called it. In the last 240 years since the Industrial Revolution started in 1771, says Perez, the world has seen the following great "surges of development" (Perez, 2004):

- § Age of Steam, Coal, Iron and Railways (1829);
- § Age of Steel and Heavy Engineering - electrical, chemical, civil, naval (1875);
- § Age of Automobile, Oil, Petrochemicals and Mass Production (1908);
- § Age of Information Technology and Telecommunications (1971).

The last Age is still ongoing, so that the next revolution, possibly the Age of Biotech, Bioelectronics, Nanotech and New Materials is still uncertain – both in terms of contents and beginning. Hence, the current technological revolution has not yet delivered its full potential, according to Perez, since we are still in the middle of the surge associated with changes brought by the Information Technology and Telecommunications Age.

In Perez's theory each surge of development is broken into two periods: the first is dominated by finance and marked by speculation and financial bubbles; the second is dominated by periods of production and "deployment" of innovation, marked by prosperity and security, and generally regarded as a "golden age" (Perez, 2010.) The transition between the finance and the production phases is not a smooth one, but it is generally marked by panics, recessions, and political confrontations - as in the times we are living today.

What are we to do then to realize the full potential of the current Age of Information Technology and Telecommunications? First, we have to move the economy away from the biased perspective of short-term goals pursued by the stock markets in the transition phases and drive it toward a culture of long-term investments. To achieve this goal, Perez (with a growing number of other voices) asks for heavier state involvement and stronger government action. A preliminary consequence is that in a globalized world this entails globalized regulation – at least at a basic level, and particularly in those sectors with no physical frontiers, such as finance, the environment and telecommunications. Local regulations also play a significant role because of the changes affecting many businesses which, by leveraging on globalization, are identifying the best locations to maximize productivity and profits. This is causing, particularly in the high-tech industry, a huge change, with a shift from hierarchical, authoritarian structures of the past towards more lateral, consultative "network organizations" in the present. Of course, even local, regional and national governments have to respond quickly to adapt their own organizations to the described changes in businesses – if they want to succeed in building societal consensus on the kind of investments they want to attract.

These are clear signs of a shift in the business climate as described by Perez: to help society embrace a new, more sustainable development paradigm – with a reduced dependence on high-carbon industries and products, and without returning to centralized, national endeavors – new ways of collaborations must be found. Found by exploiting the clear advantage of "networked organizations" between heterogeneous players, involved in pilot projects launched by newly resurgent states – decentralized, multi-level, innovative, experimental and, above all, consensus building – and replicated either nationally or locally.

In these new forms of collaboration, governments, in particular, are not limited to regulation, but play a more active role "in bulding things". And if in the past technology revolutions of new infrastructures ranged from canals to electricity, from railroads to telephony, today "the revolutionary thing" is *Broadband Internet*. This is the core of today's problem, according to Perez: broadband, low cost Internet access for all, is the "social and geographic frontier of the global market", "the most gigantic space for market action", "a dynamic demand space" where companies should focus their actions, while managing the paradigm shift in organization they are experiencing – with people getting used to work more flexibly and with greater, specific emphasis on innovation (Hegedus, 2009.)

And even if governments must help create this space, help people get connected, help organizations leverage on their paradigm shift, the only way to move from an economy based on zero-sum confrontations to positive-sum confrontations is to abandon "top-down approaches forcing separate interests to be reactive and self-centered", in favor of a common strategic vision developed jointly by the government, businesses and the society. Corporations, in particular, must change the way they do business. The return to "a golden age" is centered on a new concept of business leadership focused on corporate social responsibility and trials of new business models. Business executives are being called on even more to become social leaders.

In effect, since Internet is a transparent world, there are fewer chances to hide: a more sustainable approach to business, inspired by *coopetition* (see, for example, Dagnino and Padula, 2002) more than competition, a good business reputation and a good name can do wonders for a brand. If a corporation behaves like a good citizen it will be easy to put in practice its formulated vision - together with the government and other partners. *Broadband Internet* in the current Age of Information Technology and Telecommunications, can be the crucial factor to move the economy and the society out of the transitional phase of the current technological revolution, thus leading to a real knowledge society where many demographic, health and climate change problems will have been solved.

In this paper we will show that the realization of complex programs to expand business – particularly those enabling the evolution towards a smarter model of society and based upon cooperation among universities, ICT industry, public administrations, and SMEs - requires first this described change in the way of doing business. Companies that do not understand this and do not adapt will not be positioned to take advantage of new, coming, business opportunities, driven by the deployment of the innovations brought by the Broadband Internet revolution. In the following sections, we describe first the need for a model that involves a staged collaborative approach in the cooperation between ICT industry, universities, PA, and SMEs, and then in detail the approach we have developed.

2 THE NEED FOR A NEW COLLABORATIVE APPROACH

As outlined above we are now at a "turning point" where everyone, and with urgency, is requested to understand our economy's weak points and to overcome them with the courage to change. Business valuations for investment decisions often rely on financial issues not including intangibles and qualitative information. The lack of environmental or social performance indicators does not help in this way. This concern is the object of discussion now by some market participants – including regulators, business associations, analysts and investors – who think that investment decisions and business valuations could be enhanced if they properly reflect environmental, social and governance risks that often tag along with them. In the European context, UNEP FI (The United Nations Environment Programme Finance Initiative) works to develop and promote linkages between the environment, sustainability and financial performance.

Market actors need to make the effort to clearly understand society's needs and to implement an innovative way of thinking about the organization, the internal and external process and the relationship with customers, suppliers, society and environment. Thinking about a business in a "smart" way implies sustainability being placed at the epicenter of strategic vision and business models and integrated into strategic planning, R&D, go-to-market strategy and investment decision-making.

The complexity of today's global market requires a big effort in targeting value, mitigating risk, optimizing capabilities, and aligning the organization to achieve objectives. Unfortunately, companies are often pushing innovations relating to the longer-term future down in priority, to avoid becoming vulnerable, since these innovations don't optimise earnings. Instead, the current situation needs a "smarter collaborative approach".

2.1 Do Business in a "Smarter Way"

"Smart business" based on *coopetition* between corporations, governments, universities and other players must lead the way by turning new challenges into opportunities. In the past, technological innovation has consistently contributed to growth and, according to Dearing (Dearing, 2000) sustainable development offers today an organising framework for business to grow, based on opportunity and respect for human values. We expect that sustainable innovation will allow economic and social growth of cities and societies without harming the environment thanks to:

- a wide variety of new strategic opportunities, which enhance the economic and societal value;
- significant investments in new sustainable technology and products, and innovative services that employ high-tech developments in telecommunications and information processing;
- a push from governments encouraging experimentation that fosters innovation and improves sustainability.

And even if this is true for new cities, where the digital revolution is deeply influencing the way they are designed, first impacts are seen also on existing cities - particularly because of new advanced models being developed to help city planners to agree on funding priorities for new infrastructures and new services (see Andreoli and Medaglia, 2010). In effect, the last few years have seen, across the globe, an increasing focus on cities where the rapid growth in population poses new challenges for city services and infrastructure, but at the same time, creates new economic opportunities and social benefits for people through the offering of critical infrastructure and services, from public transportation to healthcare and education.

In the city, institutions have to take into consideration the problems coming from this rapid urbanization, the scarcity of resources and the increasing need for technology infrastructure. The promise of the Smart City vision is to ensure that the economic and social benefits of urbanization are actually realized by curbing its negative consequences, thus allowing people living a smart life to be integrated in the network with all devices connected. The benefit is an increased quality of life, as technology allows saving time that will be devoted to private life - while increased efficiency reduces CO2 emissions. Some cities have already started "ad hoc" actions by revamping their critical infrastructure and services through dedicated projects. There are also other cities that are incorporating the Smart City vision in its entirety in the strategic plan of the city. Two are the aspects on which a great focus is needed in a Smart City project:

- technological, related to the smarter use of the ICT technologies;
- strategic, related to the strategic thinking of the communities and its leader..

From the technological point of view, what makes a city "smart" are the combined use of software systems, server infrastructure, broadband network infrastructure and client devices to better connect home, public administrations, education, healthcare, public safety, real estate, transportation, and utilities.

From the strategic point of view, stakeholders need to implement a "long term strategic thinking", collaborating and planning to grow in a way that ensures the city stays healthy on all levels. In a globalized world without environmental and telecommunications frontiers, the Smart City vision is extended to the complete society. But the complexity of the task and these goals require a change in the way of working of firms, institutions, and universities, pushing them to cooperate on projects that bring high empowerment to individuals and societies and to engage key stakeholders, in order to drive the complete process and to realize projects where long term objectives other than short term profits are taken into consideration.

2.2 From Short Term to Long Term Strategy

Short-term pressures often tend to win against long term interests and time constraints. An excessive short-term focus can discourage long-term value creation and investments. The business model driven by shareholder value objectives has been highly successful for businesses in the past; however, driving businesses by investment decisions based on quantitative analysis and short-term financial performance evaluation is now unsustainable. When the performance of a company is based only on financial return, as done in the traditional business model, the resource consumption and the lack of value for stakeholders is not taken into account. The complexity of today's challenges puts new requirements on the business approach:

§ a change of perspective from shareholder to stakeholder value generation;

§ the correct importance given to the company stock value, that should never be privileged with respect of the future development of the company, sacrificing long term investments like R&D and innovation;

§ a clear relationship among the objectives of industry development, care of the environment and social benefit;

§ careful consideration before implementing any action (such as moving the enterprise line to reduce costs) of how it impacts global and local societal trends and the environment;

§ evaluation of the company performance in a different way.

Given the social implications of these projects, carried on to generate value for a "multiple set of stakeholders", the related "Smarter Business Model" can not be based on anything different from real cooperation between institutions, the industry and universities with full commitment to achieve the common project goals. Working collaboratively is something that we often talk about but has been difficult to put into practice. In reality, this is a philosophy, perspective or way of working together with a clear focus on the agreed long term goals and value generation for all the stakeholders - based on mutual trust and respect. Collaboration furnishes the company, from strategy development through implementation, the ability to transform and overcome resistance to change.

3 THE STAGED COLLABORATIVE APPROACH

What we have attempted to demonstrate in the above sections is the need for a new approach to doing business today during this "ICT-everywhere" revolution. Not only are new models of business required but new means are demanded for defining and generating business models. Our proposal concerns exactly these new means: a new approach to defining successful business models in our rapidly evolving ICT economy.

This approach encompasses three essential elements. The first concerns an innovative model for cooperation between different stakeholders: enterprises, universities, and municipal entities. In the past such cooperation has been attempted but with limited success, at times achieving some narrowly defined short-term goals, but rarely having a sustainable impact on the improvement of the community as a whole. The model for cooperation proposed in this paper is a *collaborative* one. Collaboration means aligning the goals among the participants towards a sustainable result. Cooperation in a collaborative manner implies that the partners don't have disparate individual goals where cooperation is a means to tie together these goals, but instead the collaborative project defines from the start goals which are *in common* among the partners. In particular these goals will be aimed at an improvement of the community that will be to the benefit of all, including, not least, the citizens themselves.

The second essential element concerns the means for managing the collaboration towards guaranteeing the achievement of these goals. This involves a staged approach with well-defined results at each stage. Once again, the idea of staging work towards defined goals is not a new one in itself. Instead, our approach is innovative in the way this staged approach is developed. Our staged approach contains, as fundamental parts, those of identifying, defining, managing, and assuring *sustainable results*. Stages involve not just defining a work plan, but identifying projects which meet fundamental objectives, defining results, assuring these projects achieve defined results, and generating business models towards the sustainability of the results.

The third element of our approach concerns the fundamental objective that brings the partners together, convinces them of the importance of such approaches, and as such acts as the motivational glue for this work. This element is contained in the title of this paper, *Smart*. "Smarter business models for smart living, smart companies and smarter societies" means leveraging ICT to the end of furnishing substantial and sustainable gains in the quality of life for people, companies and society as a whole, in a *positive-sum game*.

3.1 Smart Living, Smart Companies and Smarter Societies

What do we mean by smart living, smart companies and smarter societies, and how will these be brought to fruition by smarter business models? Smart living is that particular benefit of these business models that brings together the traditional meaning of "quality of life", intended as a more affluent lifestyle, with the progressive meaning of a healthier lifestyle with lesser impact on the environment and reduced waste of resources. Smarter business models are those particular ways of doing business that give the lie to the common belief that these two concepts are in conflict. These smarter business models emerge from the application of our collaborative staged approach.

Smart companies are those companies that will adopt these smarter business models if they want to be part of and flourish in this revolution. Smart companies will realize that the *collaborative* approach that we propose in this paper is the only way to prosper in this changing economy where the old way of doing business by a blind pursuit of myopic goals will be penalized by smarter societies.

It is evident that, as we have stressed in the previous sections of this paper, societies of the future will be smarter; smarter in the sense that their choices for development will be carefully aligned via coordination among stakeholders in order to permit sustainable growth. Smarter societies will choose to favor businesses whose work increases the *real* quality of life for its citizens. Smarter societies will insist that sustainability is essential to the livelihood of their municipalities. And we claim that the surest and fastest path to this goal is through the adoption of the staged collaborative approach we propose.

3.2 The Staged Approach towards Smarter Business Models

We propose, then, a *staged* collaborative approach towards the definition of smarter business models organized in four phases:

§ Identify the needs;
§ Define the actions;
§ Trial the ideas;
§ Define the business model(s).

We now describe in more detail each of these phases. In the first phase an in-depth analysis of the current situation in the region of interest is carried out to identify the needs of the community (this in contrast to the "usual" approach of defining projects where the starting point is a pre-identified need.) During this phase - once clearly identified the needs - metrics by which we can measure the success of projects designed to satisfy these needs, are defined. We call these *smart metrics*. We have more to say about these below.

In the second phase of the approach the means to jumpstart smart technologies which address these needs in an optimized manner with respect to current financial constraints, and with respect to the overall benefits of the stakeholders, are defined. A series of measures, in the form of projects, are proposed at this point. At the same time, the results which are expected, in line with the smart metrics defined earlier, are explicated for each proposed project, describing quantitatively and qualitatively how the projects respond to the community needs.

In the next phase, the trial phase, the previously defined projects are put to test. For each trial the results are measured using the smart metrics defined in the first phase and compared to the expected results stated for the project.

By taking these results, analysing them, and judging their success, in the last phase, business models – smarter business models – are generated for bringing these ideas to fruition in the community.

Critical to our approach, and its commitment to sustainability, is the definition of the smart metrics. These metrics are defined in such a way as to not guarantee simply an immediate and temporary solution to a need, but rather to favour the creation of a virtuous cycle to generate new business and opportunities in the region. As such the business models generated in the end of this approach do not signify a one-shot solution, but rather the beginning of beneficial business for the society.

Finally, concerning the organization of the collaboration, although each entity - enterprise, university, and municipality - is a stakeholder, each has a well-defined role. Universities naturally hold the leading role in the first phase, generally having the experience and resources adapted to the investigative and analytic work needed, but working closely with municipalities. Enterprises, given their technological role and experience in project management, will tend to lead the work in the second phase and third phases, but clearly working closely with university and municipalities. And the last phase of defining the business models will see all stakeholders sitting at a round table, each giving its appropriate input to the main objective of the process.

4 CONCLUSIONS

We have proposed a new, innovative approach for collaborative projects involving business, universities and municipal entities. We have found motivation for our approach in an analysis of the current economic situation observing that current financially-driven investment models will have ever more difficulty succeeding in the future, and instead the way towards successful business models is via new *collaborative* approaches. Although such collaborations have been tried in the past, success in projects with such diverse partners has often not been fully achieved. Our approach includes several new aspects which differentiates it from others:

§ it takes into consideration the larger benefits of projects to the society as a whole, which benefits bring sustained success to all the participants in the collaboration;

§ a virtuous cycle whereby the business models output and the projects resulting from the collaboration have as a fundamental element continuous sustainable development;

§ taking as a starting point important needs of the society and incorporating well-defined metrics to evaluate proposals.

We strongly believe that companies which continue doing business as usual, and which do not adopt new business models like those generated from our approach will not succeed in the future economy. Instead following our approach will bring real sustained benefits to all, resulting in a *win-win-win* situation.

Concerning future work, we are planning to build on the work presented in this paper in several ways. We will be proposing the creation of a nationwide entity to oversee the kind of “Smart Projects” generated by our approach, in order to promulgate this approach and provide guidance for those wishing to apply it. We will also investigate how current industrial associations can play a role in our collaborative approach. We are also working on a report demonstrating the social and demographic consequences for those businesses and communities which don’t insist on following implementing smart business models.

REFERENCES

Andreoli, G., Medaglia, C.M., 2010, Planning for a smarter society. In *Ericsson Business Review, # 1 2010.*

Dagnino, G. B., Padula, G., 2002, Coopetition Strategy: A New Kind of Interfirm Dynamics for Value Creation. Presented at *EURAM – The European Academy of Management Second Annual Conference - “Innovative Research in Management” Stockholm, 9-11 May 2002.*

Dearing, A., 2000, Sustainable Innovation: Driver and Barriers. In *Proceedings of the OECD Workshop on Innovation and the Environment, June 2000.*

Hegedus, N., 2009, Riding the Waves and Understanding What Hit Us. In *Ericsson Business Review, #3 2009.*

Perez, C., 2002, *Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages*. Edward Elgar. Cheltenham, U.K.

Perez, C., 2004, Finance and technical change: A Long-Term View. In *The Elgar Companion to Neo-Schumpeterian Economics*, H. Hanusch and A. Pyka, eds, Edwar Elgar. Cheltenam, U.K.

Perez, C., 2010, The Financial Crisis and the Future of Innovation: A view of Technical Change with the Aid of History. *Working Papers in Technology Governance and Economic Dynamics #28*. Tallinn University of Technology, Tallinn and The Other Canon Foundation, Norway.

Schumpeter, J.A., 1911, *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle.* Translated to English by Redvers Opie, 1934. Harvard University Press. Cambridge, Mass.

SARDANA: AN ALL-OPTICAL ACCESS-METRO WDM/TDM-PON

J. Prat[1]1, J. Segarra[1]2, V. Sales[1]3, J. A. Lázaro[1]4, P. Chanclou[2], R. Soila[3], S. Spirou[4], A. Teixeira[5], G. M. Tosi-Beleffi[6], I. Tomkos[7]

[1]Universitat Politècnica de Catalunya (UPC), J. Girona D5, 08034 Barcelona (Spain). [2]OrangeLabs (Lannion), [3]Tellabs (Espoo), [4]INTRACOM Telecom (Athens), [5]IT (Aveiro), [6]ISCOM (Rome), [7]AIT (Athens)

1jprat@tsc.upc.edu, 2jsegarra@tsc.upc.edu , 3vicent@tsc.upc.edu, 4jose.lazaro@tsc.upc.edu

Keywords: access networks, Fiber-To-The-Home (FTTH), next generation Passive Optical Network (ngPON), Wavelength Division Multiplexing (WDM).

Abstract: A new optical access network, named “Scalable Advanced Ring-based passive Dense Access Network Architecture” (SARDANA), is presented. It transparently integrates WDM metro and TDM PON access technologies, implementing ring protection, 100 km reach and up to 1024 users served at 10 Gb/s, with passive highly-shared infrastructure. The introduced innovations are hybrid ring/tree WDM/TDM Passive Optical Network (PON) architecture; a resilient remote node (RN), which is distantly pumped from the Optical Line Terminal (OLT); and a reflective ONU (Optical Network Unit); as well as an enhanced Medium Access Control (MAC) protocol.

INTRODUCTION

Fiber-To-The-Home (FTTH) technologies and new network architectures have to enable universal communication with one order of magnitude increase in terms of connected users, capacity and distance reach, as well as incorporate enhanced security, scalability, service integration and other key functionalities. These are the goals of the FP7 European research project “Scalable Advanced Ring-based passive Dense Access Network Architecture” (SARDANA), undertaken by seven partners in 2008-2010 [1], towards the construction of a future-proof access-metro optically converged network that also minimizes the infrastructure and maintenance requirements while keeping compatibility and integration with existing standards. Since operators face a high degree of uncertainty at this level (take rates, user demands, extension branches, etc.) and the necessity of deferring the investments, incremental scalability is a major objective.

SARDANA aims at serving more than 1000 users spread along distances up to 100 km, at 10Gbit/s, with 100M-10Gbit/s per user in a flexible way, transparently combining new OLT/ONU equipment with WDM metro network transmission and protection schemes in a next generation Passive Optical Network (ngPON).

SARDANA NETWORK ARCHITECTURE

In search of high scalability and trunk protection, SARDANA implements an alternative architecture of the conventional tree WDM/TDM-PON, organizing the Optical Distribution Network (ODN) as a WDM bidirectional ring and TDM access trees, interconnected by means of cascadable optical passive add&drop Remote Nodes (RNs), as is depicted in Fig. 1.

The ring+tree topology can be considered as a natural evolution, from the conventional situation where metro and access networks are connected by heterogeneous mixed optical-electro-optical (O/E/O) equipment at the interfaces between the FTTH OLTs and the metro network nodes, towards an integrated metro-access network. In this case, covering similar geographical area, users and services; but concentrating electronic equipment at a unique Central Office (CO), and implementing an all-optical passive alternative, operating as a resilient TDM over WDM overlay. Depending on the scenario, the ring+tree mixed topology optimizes the

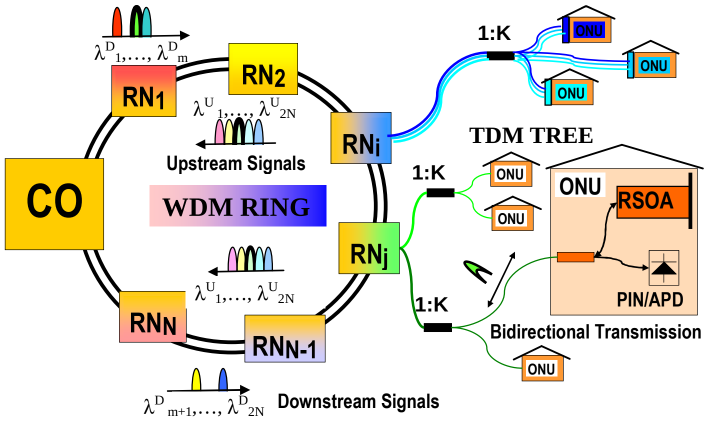


Figure 1: SARDANA network architecture with bidirectional metro WDM ring and TDM access tree.

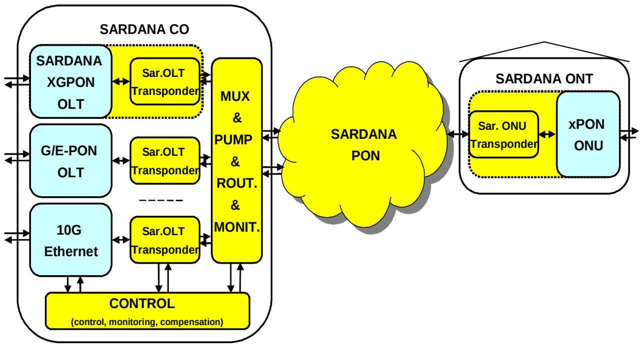


Figure 2: General network model, transparent to xPON standards.

usage of the fiber infrastructure in the ODN, and also offers enhanced scalability and flexible distribution, as new RNs can be installed.

The project mainly focuses on the physical layer, specifically on the optical and electro-optical (E/O) subsystems, being highly transparent to the protocol, coding and bit rates of existing xPON standards. The optical parameters are changed (e.g. the wavelengths band), but the chipset is kept compatible with ngPON, for a smooth migration and interoperability. The ITU-T Gigabit-capable Passive Optical Networks (GPON) standard [2] is taken as the reference and adapted to a new transparent optical layer, with WDM and active E/O devices using the existing metro/access optical passive infrastructure, as illustrated in Fig. 2.

For the full network demonstration, a down/up 10G/2.5G MAC protocol, based on FPGA, compatible with the GPON Transmission Convergence (GTC) layer has been developed in the project, and advanced new broadband multimedia services will be exhibited in the field trial.

KEY PHYSICAL SUBSYSTEMS

The implementation of the network subsystems: passive Remote Node (RN), colorless Optical Network Unit (ONU) and the Optical Line Terminal (OLT) at the CO, encompasses a number of technical challenges. Although several solutions have been investigated, the selected implementation for network demonstration is made on the basis of cost and robustness, leaving more complex advanced solutions for parallel research (Figs. 1 and 3).

Remote Node

The RN is a key element of the SARDANA network, and many of the performances and functionalities of the network depend on its design, like protection and routing. It implements cascadable 4-to-1 fiber optical add&drop function, by means of athermal fixed filters, splitters that perform spatial diversity for protection and distribute different wavelengths to each of the access trees, and remote amplification, introduced at the RN by means of Erbium Doped Fibers (EDFs) to compensate add&drop losses. Optical pump for the remote amplification is obtained by pump lasers located at the CO, also providing extra Raman gain along the ring.

This new network element, passive but with dynamic behavior, incorporated in the new PON, is not present in current standards, but it inherits concepts from the following existing standards:

- ITU-T G.984.6 on PON Extender Box;
- ITU-T G.973 on remotely pumped amplifier (ROPA) for submarine systems;
- ITU-T G.983 PON protection;
- ITU-T G.808 Generic protection switching.

The RN encompasses some key challenges, like passiveness in the sense of not using electrical supply, efficient 1480 nm pump use, and burst mode amplification generating gain transients. The inset in Fig. 3 shows a RN with 2 drop wavelengths (2 trees) and bidirectional remote amplification in the drop. Wavelength extraction is done by means of two athermal thin-film Optical Add Drop Multiplexers (OADMs) at alternated 100 GHz or 50 GHz ITU-T grid channels. The natural gain transients due to the amplification of dynamic burst-modes of the PON upstream are cancelled in this RN thanks to the crossed wavelength direction design and co-amplification of higher power continuous downstream, also avoiding Rayleigh backscattering of the ROPA. The implemented RN presents 1 dB insertion loss in by-pass, 6 dB in drop/add, and >30 dB rejection. The losses are largely compensated by about 14 dB gain of the EDF. In [3] this is compared to other types of Extender Boxes for PONs, in terms of reachable trunk&access power budget. We specify up to 16 RNs; thus, employing 32 wavelength channels, with a splitting ratio between 1 and 32 each, services up to 1024 ONUs.

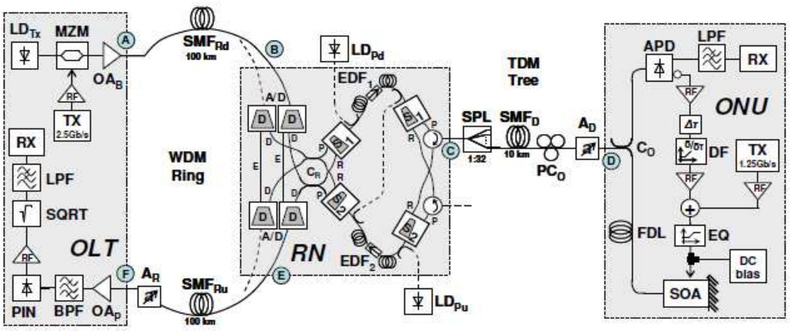


Figure 3: OLT, RN and reflective-ONU demonstration test-bed.

Lately, a reconfigurable RN has been also assembled, operated with optical power by means of particular power converting/harvesting modules, controlling latched optical switches or tunable power splitters, that enter into play at network protection and balancing [4].

Optical Network Unit

A key requirement of the ONU is to be colorless and to reuse the down wavelength, in full-duplex operation compatible with xPON electronics. A reflective-ONU optical transceiver based on Reflective Semiconductor Optical Amplifier (RSOA) has been taken as preferred option because it is the cheapest available choice for the WDM-PON, although it can rise up serious impairments operating in full-duplex with wavelength reuse. To overcome the bandwidth, noise and crosstalk limitations, a complete study of the possible optical modulation formats has been done and several compensating techniques have been developed:

- reduced Extinction Ratio (ER) downstream with feed-forward cancellation at ONU [5];
- wavelength dithering to reduce Rayleigh backscattering and reflections [6];
- upstream chirped-managed RSOA with offset-filtering, reaching 10G operation [7];
- adaptive electronic equalization, using MLSE and DFE/FFE at 10G [7,8];
- integrated colorless optical FSK demodulation with a SOA/REAM [9];
- wavelength shifting at ONU for reduction of Rayleigh scattering [10];
- other modulation formats tested like SCM, SSB and homodyne PSK are kept as longer term research.

For example, Fig. 3 shows the WDM/TDMPON system scheme implemented with downstream cancellation based on feed-forward injection and square root equalization, employing intensity modulation and wavelength reuse over 100 km reach and high split. An optimum downstream ER of about 3 dB was deduced for the given budget and receiver sensitivities. For the FEC threshold, the ER can be increased, maintaining down- and upstream. With an ER of 4 dB, the upstream penalty is 2.9 dB while the downstream benefits from 1.2 dB over an ER of 3 dB (Fig. 4).

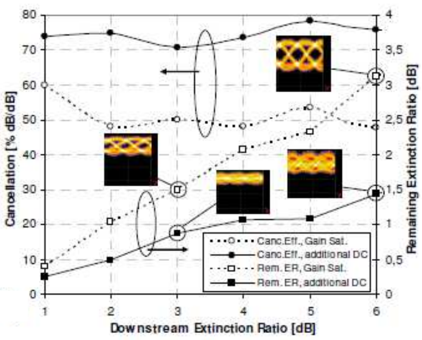


Figure 4: Downstream Extinction Ratio cancellation measurementS.

Central Office

The CO furnishes the light generation for the whole network and its control. Optics at the OLT includes: WDM multiplexers, optical pre/post-amplifiers, equalizers, and protection switches and monitors.

NETWORK TESTS

First tests of the SARDANA network have been performed, in different configurations. Fig. 5 shows the scheme and results in a 105 km ring between Rome and Pomezia cities, at 10G down- and 2.5G upstream, with 2 RNs and 3 channels; the pump power was below 1.2 watts at 1480 nm. Sensitivities are -33 and -36 dBm respectively. Protection against fiber cut was validated, with less than 1 dB penalty at rerouting, in down- and upstream directions.

With the burst mode upstream operation, any gain transients at CO or RNs EDFs can be mitigated by means of pre-distortion carving of data packets at the ONU, allowing a strong reduction, to 30%, of the packet overshoot [11]. On the other hand, and because of the highly variant optical traffic at the ring, it is useful to develop an automatic method based on a genetic algorithm to assess and minimize the impact of nonlinear crosstalk in WDM ring channels; by optimizing channel frequencies and powers, the budget is improved in 3-5 dBs [12].

PROPOSED MAC PROTOCOL

Considering the increased bandwidth, number of users and distances, as well as the interactive services that SARDANA plans to demonstrate, the design of an advanced new MAC protocol plays an important role. GPON [2] is the access standard being deployed by many operators and is taken as reference in SARDANA, but the Dynamic Bandwidth Allocation (DBA) algorithm for multi-service Quality of Service (QoS) is not specified and its implementation is open; hence, to provide QoS, a fair DBA adapted to bursty traffic is proposed. The DBA is validated by traffic simulations.

GPON Frames and MAC with QoS

GPON provides various transmission rates in both downstream and upstream directions. In downstream the bit rate can be either 2.5 Gb/s or 1.25 Gb/s, whereas in upstream the rate can be selected from 622 Mb/s, 1.25 Gb/s or 2.5 Gb/s. Rates up to 10 Gb/s are under consideration in the 10G GPON study by FSAN consortium [13].

The GPON frame duration is 125 µs for both down and up-transmission rates, supported over the GPON GTC layer [2]; as shown in Fig. 6.

GPON transports Ethernet or IP frames using the GPON Encapsulation Method (GEM). This enables fragmentation, encapsulation and extraction of variable client frames with different traffic types to support diverse QoS requirements, allowing efficient transport in GEM packets (GEMs). The GEM contains the GEM header, with information to address the ONU, and a payload up to 4095 bytes. The GEMs are allocated in the GTC payload. Downstream Header includes the BWmap fields, which specify the granted data queues of the ONUs, identified by the Allocation Identifier (Alloc-ID).

In response to the BWmap granted allocations, upstream GTC is composed of a number of transmissions bursts coming from the ONUs. Each upstream burst contains a Physical Layer Overhead (PLOu) and one or more bandwidth allocation intervals, associated with individual Alloc-IDs, which contain a Dynamic Bandwidth Report upstream (DBRu), specifying the amount of data buffered in the ONU corresponding to this Alloc-ID, and the payload where the GEMs are allocated. The traffic arrived at the OLT and at each ONU is classified in separated QoS classes of service (CoS) and placed in corresponding

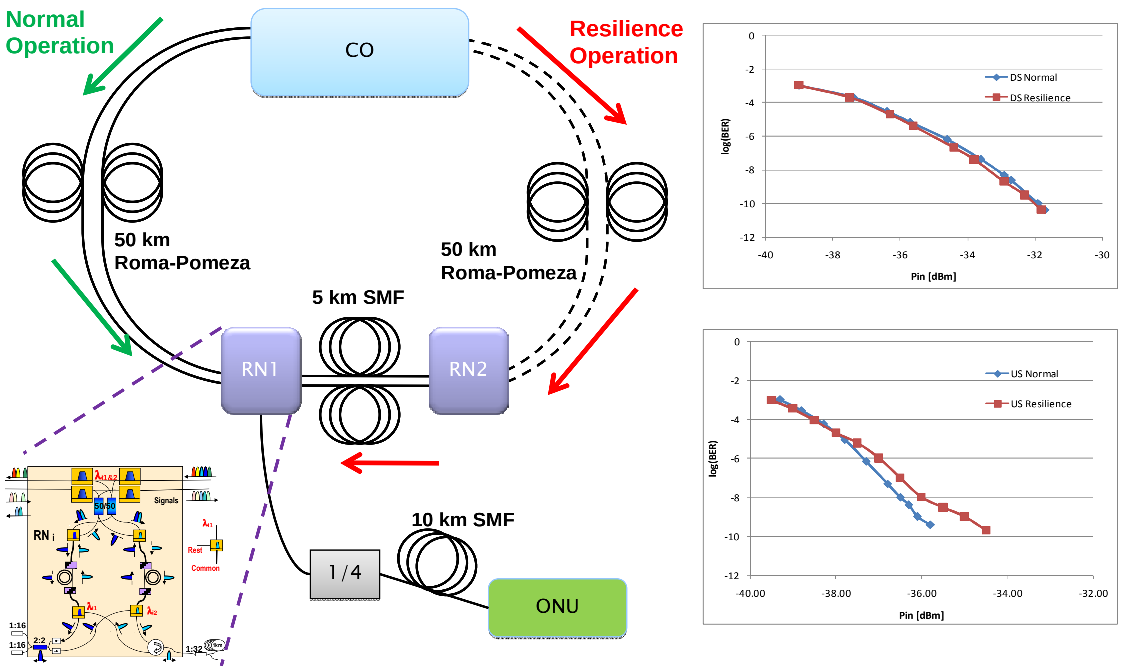


Figure 5: Network test configuration with 105 km ring, RN scheme (inset) and down- (above) / upstream (below) transmission BER measurements.

queues Traffic Containers (T-CONTs), so that it can be treated in a different manner by MAC protocol run at the OLT.
According to G.983.4, traffic is classified in the T-CONT CoS types 1, 2, 3 and 4:

- T-CONT type 1. Fixed bandwidth is supported, like emulation of leased line services and Constant Bit Rate (CBR) applications. Corresponds to the SARDANA Premium Class, using a permanent portion of bandwidth matching its Committed Information Rate (CIR).
- T-CONT type 2. Supports assured bandwidth for Variable Bit Rate (VBR) traffic, with both delay and throughput requirements, such as voice and video. Corresponds to the SARDANA Silver Class, but there having an assured CIR bandwidth plus an extra Excess Information Rate (EIR) assigned in DBA manner.
- T-CONT type 3. Better than best-effort services offering a guaranteed minimum rate. Corresponds to the SARDANA Bronze Class, with guaranteed rate up to its CIR and surplus EIR bandwidth granted by DBA.
- T-CONT type 4. Best-effort services, such as browsing and FTP, up to a maximum rate Rmax and receiving bandwidth as the higher priority types do not use it. Corresponds to the SARDANA Standard Class using any bandwidth left.

The scheduling discipline applied in downstream is very simple: a Strict Priority (SP) data burst queuing discipline is employed, serving the data bursts in order of their priority and considering also the particular Service Level Agreement (SLA). Bursts with higher priority are attended first and bursts with lower priority wait to be served until there are no bursts with higher priority to be considered.
In upstream the polling Status Reporting (SR) technique [2] is used in order that the

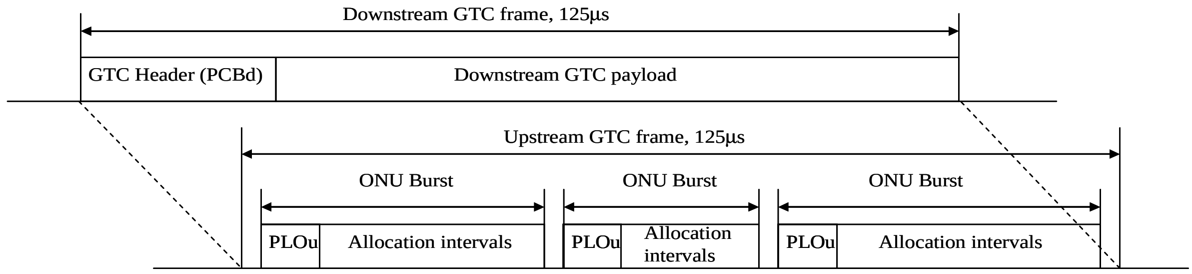


Figure 6: Downstream and upstream GPON Transmission Convergence (GTC) layer frame formats.

OLT knows the T-CONT bursts generated and stored in the ONUs, and subsequently the OLT will create the proper grants to the ONUs, which will send its authorized data bursts in the upstream frame (Fig. 6). The SR, informing the burst length of existing T-CONT bursts in the ONU in a nonlinear code, is furnished every Scheduling Interval (*SI*), which is the number of GPON frames of the ONU polling cycle. The *SI* is a critical parameter and must be adjusted to the traffic and propagation conditions, so that the OLT achieves global and current information on time about the T-CONT bursts generated in each of the ONUs. Once the OLT gets knowledge of the ONUs traffic needs, it will distribute the available bandwidth, applying the SP discipline (with a round-robin in case of several bursts in queue of the same CoS) plus particular SLA, between the requests [14]. Grants are generated and inserted in the BWmap fields in the downstream frames. Moreover, when an upstream T-CONT burst is sent, additional piggybacked information about new bursts generated from the same T-CONT is also added beside the up-burst in the DBRu field, informing if a new burst of the same CoS is originated in the ONU, and the OLT reaches an updated awareness of the T-CONT bursts generation status.

The upstream operation is collision-free because all ONUs are timed by using a ranging procedure during activation and registration; thus, an extra delay may be forced at the ONU side and the Round Trip Time (RTT), which is two times the propagation time OLT-ONU, is seen as fixed and common for all OLT-ONUs pair.

Traffic Performance Evaluation

We model each k CoS data source with a VBR traffic flow, with k = 2, 3, 4. The DBA will consider the data bursts to allocate once they are generated in each T-CONT in the Optical Burst Switching (OBS) manner: a burst will be scheduled when a time edge t_{edge} is reached, with a maximum burst size B allowed [15]. The VBR data has a mean rate b_{in}; hence, the average burst length is $L_{burstk} = t_{edgek} b_{in}$. The output optical burst bit rate is b_{opt}, with $b_{opt} > b_{in}$, and the ratio $A = b_{opt} / b_{in}$ is the rate gain. The Wavelength Holding Time t_{WHTk} used for data burst transmission is

$$t_{WHTk} = t_{idle} + t_{tk} = t_{idle} + t_{edgek} / A \cong t_{edgek} / A \quad (1)$$

where $t_{tk} = L_{burstk}/b_{opt}$ is the transmission time of the data burst, and t_{idle} is a time where the reserved wavelength is idle or used by headers, being $t_{idle} \ll t_{tk}$. The optical load per active user L_{ou} in Erlangs, having every input flow the same b_{in}, does not depend on the aggregation time t_{edge}; thus, it is independent of the CoS:

$$L_{ou} = t_{WHTk} / t_{edgek} \cong 1/A << 1 \qquad (2)$$

With N_u active T-CONTs, the data burst optical load is then $A_d = N_u L_{ou} \simeq N_u / A$. The optical rate b_{opt} for downstream and upstream is taken to be 2.48832 Gb/s; thus, a rate gain A = 100 provides a mean data source b_{in} of 25 Mb/s for both down and up. In downstream the DBA is executed on real time, because the OLT is aware of all arrived data bursts. In upstream the DBA is executed every frame considering the polling and piggybacked SRs received at the OLT. The processing delays $OLT_p = ONU_p$ = 35 µs and the number of ONUs N = 32, located 20 km far from the OLT. Different traffic loads are accomplished by gradually increasing the user bit rate b_{in}. We suppose Short Range Dependence (SRD) traffic exhibition, so we generate down and up-bursts in Poisson arrivals with arrival mean t_{edge} per burst and time of service exponentially distributed (M/exp) with mean service time t_{WHT}. In a more practical self-similar scenario, we consider Long Range Dependence (LRD) burstiness behavior, which we emulate with the M/Pareto model [15]: again, the bursts are generated randomly in a Poisson distribution, but with Pareto time of service distribution with mean service time t_{WHT}. The Hurst parameter H is chosen to be 0.7, not very high because burst assembly reduces slightly the self-similarity. We consider four CoS, with traffic load distribution in 10% for the first fixed type Alloc-ID and 30% for the rest of the CoS assigned in DBA manner. The aggregation time t_{edge} is set to 10 ms for CoS types 2 to 4; then, e.g. a bit rate gain ratio A = 100 provides a mean time of service t_{WHT} = 0.1 ms. Fig. 7 depicts the downstream average delay simulation results, which include the OLT-ONU propagation time.

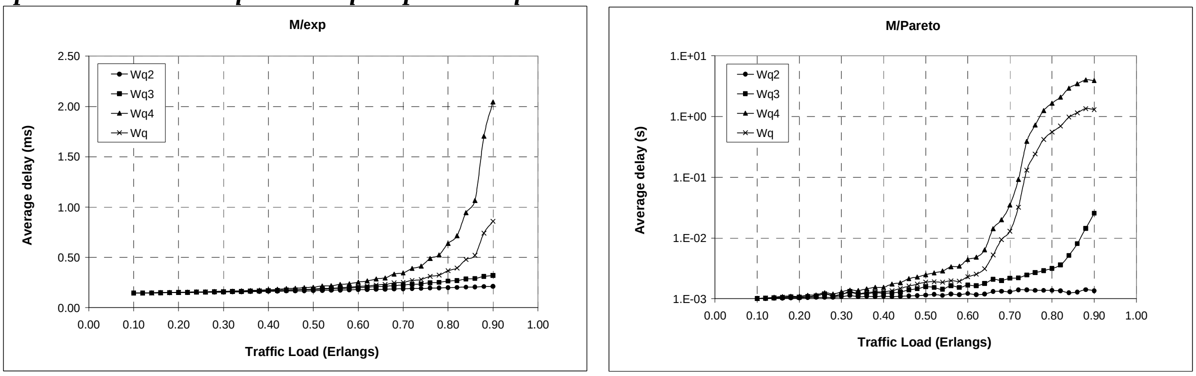


Figure 7. Downstream total average delay W_q and W_{qk} for each CoS, with Strict Priority discipline. On the left under SRD traffic, on the right under LRD traffic.

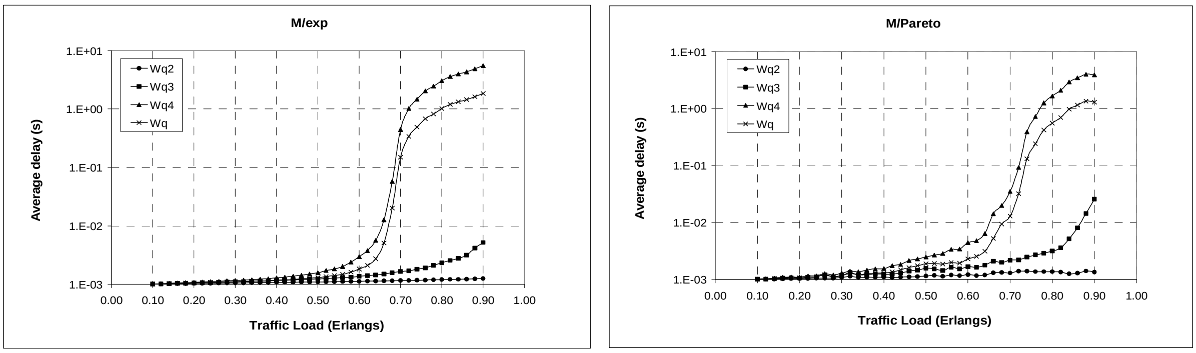


Figure 8. Upstream total average delay W_q and W_{qk} for each CoS, with Strict Priority discipline. *SI* = 6 frames. On the left under SRD traffic, on the right under LRD traffic.

The average delays obtained in downstream under SRD traffic are very low, bounded to 0.2 ms for all CoS types up to a 0.5 traffic load, confirming the correct choice of data burst aggregation time t_{edge}. The results for LRD traffic differ very much from SRD traffic because of the self-similarity, which furnishes long data bursts when transmitted; nevertheless, up to a 0.5 traffic load, the average delays are limited to 1 ms for all CoS and the latencies for CoS types 2 and 3 are kept in the order of 0.5 ms.

In upstream, to get a very good ONUs traffic knowledge on real time, the *SI* must be in the order of $t_{edge}/10$ = 1 ms (8 frames). For a 20 km OLT-ONU distance, the minimum polling SI_r is 0.27 ms (RTT + processing, 2.16 frames), which is rounded to 3 frames. The data upstream is accomplished in $2SI_r$ (traffic knowledge, BWmap report and upstream transmission); therefore, 6 frames are used for the *SI*. Simulations for both SRD and LRD behaviors provide minimum delays with *SI* = 6 frames. Fig. 8 shows the upstream average delay simulation results.

The resulting average delays in upstream under SRD traffic are bounded to 1.5 ms for all CoS types up to a traffic load of 0.5, and limited to 2.5 ms under LRD traffic and up to the same traffic load. The results for LRD traffic do not differ very much from SRD traffic because of the data segmentation in the BWmap allocation, which provides fair assignment among all data bursts to be served despite of its length. The DBA module can be improved using a bursty traffic prediction to enhance the transmission efficiency in the Long-Reach approach [14].

CONCLUSIONS

The SARDANA project for new FTTH deployment proposes a future metro-access network, involving the development of novel techniques at different levels and layers. The targeted scenario, the main new functionalities and the critical elements have been identified and the chosen solutions are tacked and experimentally validated. SARDANA network furnishes a solution for the ngPON integrating access and metro technologies, thus reaching a superior performance in security, scalability, capacity and service integration while keeping compatibility with the existing standards.

ACKNOWLEDGEMENTS

This work was supported by the European FP7 SARDANA project and the Spanish MICINN project TEC2008-01887.

REFERENCES

[1] European 7th Framework Programme project SARDANA (www.ict-sardana.eu).

[2] ITU Rec. G.984.3, Study Group 15. “Gigabit-capable Passive Optical Networks (G-PON): Transmission Convergence Layer Specification”, Geneva, Switzerland, Oct. 2003.

[3] F. Saliou, P. Chanclou, F. Laurent, N. Genay, J.A. Lazaro, F. Bonada and J. Prat. “Reach Extension Strategies for Passive Optical Networks", *J. Opt. Com. Networks*, vol. 1, no. 4, Sept. 2009.

[4] A. Baptista et al. “Improved remote node configuration for passive ring-tree architectures”, *ECOC 2008*, Brussels, Sept. 2008, paper Tu1f6.

[5] B. Schrenk, F. Bonada, M. Omella, J.A. Lazaro, J. Prat. “Enhanced Transmission in Long Reach WDM/TDM Passive Optical Networks by Means of Multiple Downstream Cancellation Techniques”, *ECOC 2009*, Vienna, Sept. 2009, paper We.8.5.4.

[6] V. Polo, B. Schrenk, F. Bonada, J. Lazaro and J. Prat. “Reduction of Rayleigh Backscattering and Reflection Effects in WDM-PONs by Optical Frequency Dithering”, *ECOC 2008*, Brussels, Sept. 2008, paper P421.

[7] M. Omella, I. Papagiannakis, B. Schrenk, D. Klonidis, A. N. Birbas J. Kikidis, J. Prat and I. Tomkos. “Full-Duplex Bidirectional Transmission at 10 Gbps in WDM PONs with RSOA-based ONU using Offset Optical Filtering and Electronic Equalization”, *OFC 2009*, San Diego, Mar. 2009, Paper OThA7.

[8] D. Torrientes, P. Chanclou, F. Laurent, S. Tsyier, Y. Chang, B. Charbonnier and C. Kazmierski. “10Gbit/s for Next Generation PON with Electronic Equalization using Un-cooled 1.55µm Directly Modulated Laser”, *ECOC 2009*, Vienna, Sept. 2009, PDP Th.3.5.

[9] B. Schrenk, J.A. Lazaro, C. Kazmierski, J. Prat. “Colourless FSK/ASK Optical Network Unit Based on a Fabry Pérot Type SOA/REAM for Symmetrical 10 Gb/s WDM-PONs”, *ECOC 2009*, Vienna, Sept. 2009, paper We.7.5.6.

[10] M. Omella, J.A. Lázaro, J. Prat. “Driving Requirements for Wavelength Shifting in Colorless ONU with Dual-Arm Modulator”, *IEEE/OSA J. Lightwave Technol.*, vol. 27, no. 17, Sept. 2009.

[11] F. Bonada, B. Schrenk , J. A. Lázaro, V. Polo, P. Chanclou, J. Prat. “Gain Transient Mitigation in Remote Erbium Doped Fibre Amplifiers by Burst Packet Carving at the ONU for Extended Power Budget PONs”, *ECOC 2009*, Vienna, Sept. 2009, paper We.P6.24.

[12] J. Reis, B. Neto, P. S. André, A. Teixeira. "WDM Ring Performance Improvement by Means of a Nonlinear Effects Crosstalk Minimization Algorithm", *OFC 2009*, San Diego, Mar. 2009, paper JThA77.

[13] FSAN (Full Service Access Network). Next Generation PON Task Group (NG-PON). http://www.fsanweb.org

[14] J. Segarra, V. Sales, J. Prat. “GPON Scheduling Disciplines under Multi-Service Bursty Traffic and Long-Reach Approach”, *ICTON 2010*, Munich, Jun. 2010.

[15] J. Segarra, V. Sales, J. Prat. “An All-Optical Access-Metro Interface for Hybrid WDM/TDM PON Based on OBS, *IEEE/OSA J. Lightwave Technol.*, vol. 25, no. 4, pp. 1002-1016, Apr. 2007.

STRUCTURED CABLING:
An Alternative to Traditional Solution in Residential Buildings

Fernando Martín-Rodríguez
Signal Theory and Communications Department. University of Vigo. ETSET, Ciudad Universitaria S/N, 36310 Vigo, Spain.
fmartin@tsc.uvigo.es

Keywords: Telecommunications infrastructures in buildings, Spanish regulations for residential buildings, structured cabling.

Abstract: This communication presents a study on the possibility of designing the communications infrastructure of a residential building using structured cabling (according to ANSI/EIA/TIA-568A). In Spain, there exist regulations that oblige to follow the traditional way (ICT norm, published in 2003). What's more, structured cabling solution can be regarded as very expensive at the first glance (for example: the distribution of television signals over IP is possible but expensive). On the other hand, structured cabling would allow resource sharing on the telephone and broadband data service that could be cost effective in the long term (this may be true especially for big buildings and residential complexes). This paper will study both solutions trying to determine rules to choose the more convenient. We will also study intermediate solutions where a part of the services (terrestrial and satellite television) is distributed in the traditional way whereas voice and data could be handled by structured cabling.

1 INTRODUCTION

1.1 Spanish ICT Regulations

In Spain, the cabling inside residential building is established by a legal regulation called ICT (Telecommunication Common Infrastructure) norm (BOE, 2003). ICT norm was first published in 1998 and updated in 2003 and has taken many benefits assuring good access to services in all new built homes.

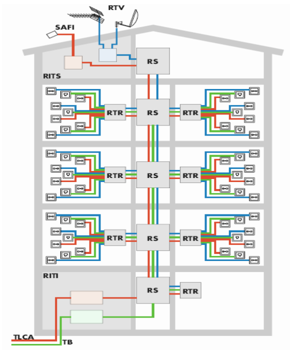


Figure 1: Typical ICT installation.

Typical ICT installations are of vertical nature (fig. 1), where terrestrial and satellite TV do up to down and telephony and broadband services go from the basement to the roof.

1.2 Spanish ICT Regulations

Structured cabling is applied in corporate buildings and ANSI/EIA/TIA-568A (ANSI/EIA/TIA, 1995) is applied as an established standard (although it is not compulsory by law).

Resulting installations (fig. 2) are structured in plants, where a horizontal cross-connect concentrates the star like cabling (generally copper) that provides voice and data service. All plants are linked to a central “equipment room” using fiber links. Using this scheme, users can share internet access (and many times they also have common resources like network servers) and also phone network access (using a traditional or IP PBX).

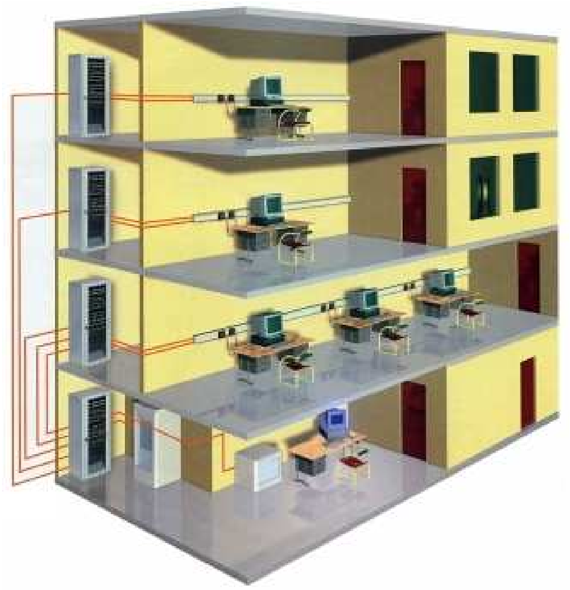


Figure 2: Structured cabling scheme.

1.3 Structured Cabling in Residential Buildings

Nowadays, there are some kinds of buildings where the services of a residential building are implemented through a structured cabling network; mainly this is true for some hotels and hospitals (where ICT norm does not apply). In this case the main problem is to implement the TV services via IP. Normally, they use equipment that converts COFDM (DTT) and satellite signals into IPTV.

In this paper we will investigate the application of this idea to residential buildings (the bigger the better) knowing that the horizontal cabling of 568A standard can be done grouping more than a single plant.

2 SERVICES AND THEIR IMPLEMENTATION

In this section we are going to summarize the typical services present in a residential building (those that are necessary according to spanish regulations) and, for each, we will comment how to implement them in the classic form and how to integrate them in a structured cabling system.

2.1 Terrestrial and Satellite TV

Distribution of terrestrial TV signals is compulsory under spanish regulations. Nowadays, most of these signals are public channels and a few are under conditional access. Nevertheless, all signals have to be present at every home.

Satellite signals are optional, but many apartment buildings distribute signals from a local operator “digital+” which are downloaded from any of two geostationary satellites: Astra 1KR and Hispasat 1C/1D.

Classical solution (as it is described in regulations) consists of a head installation (made of single channel processors, most of them amplifiers) and a coaxial distribution network (which includes signal splitters to get to every home).

To get the same service via structured cabling we need a heading made of RF to IPTV converters (EtherDVB, 2010) and we also need an IP network that is able to handle multicast traffic (that imposes a hard requirement on active equipment). See that this may be a much more expensive solution as equipment is expensive and we will need more maintenance that with a classical system. What’s more we will have a very heavy traffic added to our network. Benefits of this solution would be the possibility to establish internal channels and/or sharing pay per view material (something interesting for hotels but only feasible in very large residential communities). A possibility to relieve costs in the DVB-IP gateway could be to implement it via a PC-based solution. A PC equipped with DVB receiving cards and an IPTV server could make the trick. We could use free software like VLC for his task (VLC, 2010).

We can see that this service is difficult and less cost effective than traditional solution. So it is probably not very advisable to implement it via structured cabling.

2.2 Voice Telephony

Voice telephony is another compulsory service.

Classical solution consists of a tree-like structure of twisted copper cabling. Trunk cabling is made with multi-pair cables and on each plant we have a star-like cabling made of 1-pair or 2-pair cables. Each home should receive two pairs, although most times they will only use one of them. The origin of cabling is a room situated in the ground floor or in the basement where pairs are connected to those incoming to the building from the telephony operators. This room is called “RITI” (Recinto de Instalaciones de Telecomunicaciones Inferior, Lower Telecommunications Room).

In a structured cabling solution, we have two different options:

- Installing a PBX in the RITI. Telephone cabling would be identical to the classical one but residents would benefit from free internal calls and they would share the outward lines getting reduced costs.
- Using VoIP. Data cabling would be used for voice with small traffic cost. Telephone terminals would be more expensive but PBX would be converted into a standard PC plus software. We could use commercial software like “TalkSwitch” (TalkSwitch, 2010) or open source solutions (Asterisk, 2010).

Both solutions require an accounting system and more maintenance than a classical solution (where maintenance falls almost completely on the operator side) but for telephony we can speak about cost effective solutions for medium to large communities. In this case, perhaps the VoIP solution is the optimum.

2.2 Broadband Communications

Spanish regulations oblige building makers to leave proper cabling ducts for broadband communications. It is not specified neither which kind of cabling to use nor what services to offer. These channels are often used by cable companies to offer packed services (which usually consists of DVB-C TV and internet access). These companies also offer telephony but they normally use the “RITI” room to connect to the analog pairs that go into the homes so that for the users this service seems identical to that of the classical operators.

In this case we can become our self broadband operator and install structured cabling using the cabling ducts. Internet access would be easy to implement (and to share between residents) installing the proper equipment in the “RITI”. After that, we could implement VoIP with little extra effort. TV distribution remains as the main problem for this kind of installation. Any solution will be expensive and will add a very strong traffic to our network.

In the case of building an intra-building network we can get some more benefits as a community intranet and, perhaps, distribution of local video content. Another benefit would be the integration of security and/or home automation. It would be interesting to get an agreement with a cable or IPTV operator so that TV contents could be shared by the community.

3 DISCUSSION OF POSSIBLE SCHEMES

3.1 Data cabling Scheme

ANSI/EIA/TIA-568A standard recommends that network sockets should be grouped by plants, using what it is normally called a "horizontal cross-connect" to concentrate the horizontal cabling (made normally of copper UTP cable). That structure is suitable for corporate buildings where we have few plants and they are large. In this scheme the horizontal cross-connect must be in a dedicated room called "telecommunication room" (TR).

Residential buildings tend to be of vertical structure: many plants of reduced dimensions. Besides square meter is a very expensive resource, id est.: normally, we can not afford a TR on each plant. ICT regulations oblige to have a small common space on each plant called: "secondary register". This is merely a small closet where we can connect cables.

Figure 3: Secondary register. Regulations establish its minimum dimensions. For 5-8 homes in a plant, we have 500x700x150 (millimeters).

If we make secondary registers "relatively big" and we furnish them with electric power (something that with classical solution is almost always not necessary), we could install the necessary switches into them.

Nevertheless this is an excessive solution and we can get a better structure. So the proposed structure is:

- "Use the RITI as a telecommunication room and install inside it all the necessary switches. As this room must be connected by cabling ducts to all homes in the building, we can use them to carry the UTP cables to each user socket".

The condition to be able to implement a network in this way is that we fall within the maximum distance that ANSI/EIA/TIA-568A specifies: 90 meters. Thinking about a big building of ground floor and ten plants we could have 7 meters from the RITI to the first plant, and 27 meters more until the top plant. We would have 56 meters left for the horizontal path until reaching the furthest socket. Then we can conclude that in vertical buildings (see figure 1) we can use this solution almost always.

Sometimes we have contiguous buildings that share the RITI (sometimes it is a single big building with more than one stairwell). In this case we will still be able to use the proposed structure, as long as we do no exceed the maximum distance.

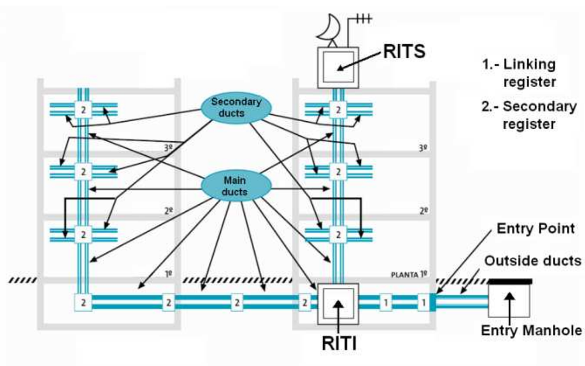


Figure 4: Double-building infrastructure.

In certain big building projects, we have more than one RITI where each serves one or two stairwells. In this case, our scheme is again valid. We should have the RITIS connected using optical fiber and one of them would provide the connection to the outside. That is exactly what ANSI/EIA/TIA-568A specifies for connecting several TR's, the main of them becomes a vertical cross-connect or ER (Equipment Room).

In vertical buildings it is very unlikely that we go over the maximum (90 m) distance for copper cabling. If we have such problem, we will have to concentrate part of the building sockets and then connect the switch with the corresponding RITI (using fiber). Perhaps the easier form is to choose a proper secondary register for this task.

In the case of big housing developments, consisting of individual houses, the assumption of short distances becomes false. This particular case is a difficult one for any kind of design. For example, in the classic TV distribution schemes, much care must be put on avoiding signal loss due to distance. In this situation, engineers must use amplifiers, normally in the secondary registers. To perform a structured cabling solution, we will have to use network switches instead of amplifiers. Note that, at least, switches will not have the "noise amplification" problem that we may have with several cascaded amplifiers.

3.2 Telephony Scheme

In the case we want to use IP telephony, nothing else must be done. The same sockets will be valid for voice or for data applications but users will have to get IP phones (or use VoIP software on their computers).

If we are to use a PBX we will have to take into account voice cabling. Nevertheless, we have two options:

- Using the classic multi-pair cables for the main ducts (trunk cabling) and single or double pair cables for distribution on each plant. With this solution, phone sockets will be RJ-11 and we will always use them for voice (with basic analog telephones).
- Using the same UTP (8-wire) cabling of data applications. We will concentrate them in the same rack of the data cables. Then we will connect data cables to switches and voice ones to the PBX (if there is more than one RITI we can use multi-pair cables to connect them, the PBX will normally be in the "main" RITI). This is a more expensive solution but it carries an advantage: we can change a socket (all of them will be RJ-45 female connectors) from voice to data or vice versa.

Note that the advantage of the second solution is important in corporate environments bur perhaps not so important in residential ones. What's more, current regulations specify as compulsory the first solution for voice cabling (we speak of voice cabling but, nowadays, operators are using it to provide data service through ADSL).

So for telephone cabling we think that in residential environments classic solution will still live for a long time (at least until the ICT norm is modified) and that future will be in the use of VoIP solutions. See that although we install voice cabling, users will still be able to use VoIP if there exists data cabling.

3.3 The Television "Problem"

As we had already commented before, television distribution is the most doubtful part of our alternative. TV distribution via IP brings few advantages and causes many problems:

- Conversion from broadcast formats to IP is expensive. Besides that has to be done to "every channel" so that it becomes even more expensive. Services with a lot of channels (like satellite digital TV) are easily handled in the classical form (where the entire frequency range is translated to an intermediate frequency and distributed together) but to convert all the channels to IP streams is nearly impossible.
- Network installation and maintenance costs are increased by the need to support multicast schemes.
- Even with multicast, network traffic is increased a lot. About a minimum of 4 Mbits/second bulk traffic (an uninterrupted stream) will be added for each channel that some user is viewing at any time.

The conclusion is that almost always it will be better to use the classical solution for the terrestrial and satellite distribution. Only in the case where pay per view service is very important (like hotels) will be useful to distribute IP video streams. And even in this case we can have both solutions working together: IP distribution for the pay per view channels and classical coaxial cabling for the public ones.

To reinforce this idea, we must comment that ICT regulations oblige to install coaxial TV distribution in every residential building. See that figure 4 shows a room called RITS. This name stands for "Recinto de Instalaciones de Telecomunicaciones Superior" (Upper Telecommunications Room). This room has to be in that position because it is holding the amplifiers that will reinforce the signals captured by the antennas. Those signals are very weak and they have to be processed as soon as possible. In the case of having IP streams converters (DVB to IP gateways); we would have to put them in the same place. Then we could connect them to the rest of the network.

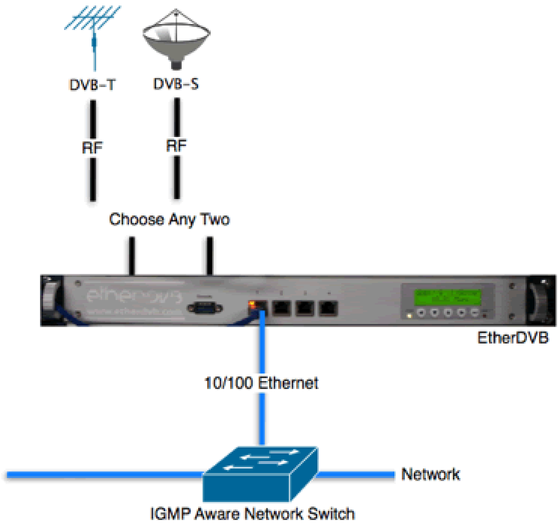


Figure 5: DVB to IP gateway.

4 LEGAL ISSUES

Taking into account spanish regulations for residential buildings the installations of structured cabling is legal as a kind of broadband cabling. The norm does not oblige to install broadband cabling (that is normally done by operators) but it is possible for the residents' community to make the installation and operation of their own system.

For the overall ICT to be legal we should install phone cabling using multi-pair cables and we also should distribute terrestrial and satellite TV via coaxial cable.

In structured cabling, the phone multi-pair cables are only useful if we are going to share a PBX (a classical one, not IP telephony). This solution could be cost effective only in large communities but probably it will be substituted by IP telephony soon.

As we have seen in the previous section, the coaxial TV distribution is, in fact, recommendable because IP TV distribution results to be very expensive (especially if we have many channels to distribute).

5 CONCLUSIONS

In this paper we have studied the possibility of extending the ANSI/EIA/TIA 568A standard (structured cabling) to residential building. That extension would permit applying the same techinques to all kinds of buildings, we would also have a full integration of services with all kinds of information flowing trough the IP network. From the user point of view this could result in new services and lower costs arising from sharing resources like internet and phone access.

One of the main conclusions is that implementation of ANSI/EIA/TIA 568A in a residential community is possible but with some limitations:

- Legal regulations (currently in effect) impose some requierements (like phone cabling) that could be not necessary in our case but that we could try to use in our benefit, for example: sharing a PBX that allows using cheap analog phones.
- TV distribution is still a difficult task using only structured cabling. The big expense of a converter for each channel and the huge increase in network traffic almost recommends maintaining a classical coaxial distribution. In this case, regulations play on our side because coaxial distribution is, nowadays, compulsory.

REFERENCES

BOE, 2003. REAL DECRETO 401/2003, de 4 de abril (ICT spanish norm). *Published in Spanish Official Bulletin (BOE) on 05/14, 2003.*

ANSI/EIA/TIA, 1995. ANSI/EIA/TIA 568A. *Standard for structured cabling from ANSI, the Electronics Industry Association and the Telecommunications Industry Association.*

EtherDVB, 2010. DVB-to-IP Gateway (data sheet): http://etherdvb.com/products/dvb-to-ip-gateway.

VLC, 2010. VideoLan (VLC): cross-platform open-source multimedia framework, player and server. http://www.videolan.org/vlc/.

TalkSwitch, 2010. IP PBX phone systems for small business (up to 64 phones). http://www.talkswitch.com/us/en/.

Asterisk, 2010. The open source telephony project: http://www.asterisk.org/.

Wavelength Division Multiplexed Passive Optical Networks (WDM-PONs) combined with wireless nodes for delivering the bandwidth for next generation access services

George Heliotis, Ph.D., George Agapiou, Ph.D.
Hellenic Telecommunications Organization (OTE S.A.), Research and Development Department,
Pelika and Spartis str., Marousi, Athens, Greece
gheliotis@oteresearch.gr,gagapiou@oteresearch.gr

Keywords: Next Generation Access Networks, Passive Optical Networks, Wavelength Division Multiplexed Passive Optical Networks, Service Provisioning, Mobile Backhauling, 4G networks, Long Term Evolution networks, WiMAX.

Abstract: Next-generation WDM-PON implementations offer the possibility of tremendously high bandwidth per customer that can satisfy even the most stringent requirements of future applications, as well as enhanced rich abilities with regard to the configuration and optimization of service provisioning. This paper presents a review of WDM-PON technologies and architectures from a network operator's perspective. The enabling characteristics of WDM-PON networks and the supported novel features of service provisioning and differentiation are examined, and a discussion of the challenges and obstacles that must be overcome for their successful implementation is provided. Apart from fixed networks, when these technologies are used in combination with wireless access nodes, they present advanced architectures for offering high bit rate services to nomadic users. Such architectures can be used to support next generation 4G networks, such as mobile WiMAX and LTE networks.

1 INTRODUCTION

We are currently witnessing a very exciting era of access telecommunications networks radical transformation. Indeed, most network operators worldwide are planning to or have already started deploying some sort of optical access, either in the form of fully optical FTTH (Fiber-To-The-Home) implementations or most usually FTTC (Fiber-To-The-Cabinet) combined with VDSL2 for the final drop to the customer's premises as an intermediate step to a fully optical access network.

Amongst the types of optical access networks (different technologies and architectures) that may be chosen for large-scale deployment, PONs (Passive Optical Networks) present a favourable solution in terms of overall cost (CAPEX and OPEX), flexibility, upgradeability and performance. The paper will present architectures for servicing customers that share the same fiber. While the supported data rates can be very high with the current PON standards, the inherent use of time division multiplexing in them poses limitations on the maximum bandwidth per customer that can be achieved and the maximum number of customers per fiber that can be supported, as well adds complexity to the operation and management of the network. To fully utilize the very high bandwidth that fibers have to offer, it is evident that wavelength division multiplexing (WDM) techniques have to be employed in the access network (as it was also done in the past in the core/backbone fiber optic networks).

These next-generation WDM-PONs offer the possibility of tremendously high bandwidth per customer that can satisfy even the most stringent requirements of future applications, as well as enhanced rich abilities with regard to the configuration and optimization of service provisioning. In this paper we present a review of WDM-PONs technologies and architectures from a network operator's perspective. Following a brief discussion of the limitations of current TDM-PONs, we examine the enabling characteristics of WDM-PON networks and the features of the multitude of novel service provisioning and differentiation that can be supported, as well as discuss the challenges and obstacles that must be overcome for their successful implementation. When the above architectures are used for transmitting high bit rate services and are used in combination with wireless antennas at the access nodes, they present advanced architectures for offering high bit rate services to nomadic users. Such architectures can be used to support next generation 4G networks, which is the case of mobile WiMAX and LTE networks.

2 Current TDM-PONs and LIMITATions

Currently the prominent PON standards are Gigabit PON (GPON) (ITU, 2004) and Ethernet PON (EPON) (IEEE, 2004), though the older Broadband PON (BPON) (ITU, 2001) is still in use in some networks but is rapidly being superseded by GPON. GPON is supported by the ITU (International Telecommunications Union), while the IEEE (Institute of Electrical and Electronic Engineers) supports the EPON.

GPON offers a generous increase in transmission speed in comparison to BPON, and can support symmetric rates of up to 2.488 Mbps, with the asymmetric version of 2.488/1.244 Mb/s (downstream/upstream) being the most usual current implementation. One of the most important advantage of GPON is that it abandons the legacy ATM encapsulation and uses the new, very flexible and efficient GEM method (GPON Encapsulation Method), which allows the simultaneous encapsulation of a combination TDM traffic and Ethernet frames. ATM traffic is still possible since a GPON frame can carry both GEM and ATM data. The most common GPON implementations use a 1:64 split ratio (i.e. 64 customers per PON branch), while the standard supports split ratio up to 1:128. The ITU is already working on the new version of GPON (the 10GPON) that will support increased downstream transmission of 10Gbps.

EPON is part of IEEE's "Ethernet in the First Mile" action is the main current competitor to GPON for modern PON-based access networks. It uses Ethernet encapsulation and supports lower symmetric rates of up to 1244 Mbps. It also shows overall lower transmission efficiency. As in the case of 10GPON, the IEEE is already working on the respective new 10GEPON standard.

It is noted that apart from their main differentiation with respect to the transmission protocol, both GPON and EPON have similar operating characteristics since they use the same simple WDM-scheme for simultaneous transmission of downstream/upstream traffic over a single fiber and they both utilise a time-division multiplexing (TDM) scheme for multiple user servicing. Specifically, the downstream traffic is broadcasted from the Optical Line Terminal (OLT) to all customers' Optical Network Terminals (ONTs) at 1490 nm and 1550 nm (video overlay). The customers' data are divided into different timeslots within the broadcast stream using a TDM scheme. In the opposite (upstream) direction, in order to avoid collisions a TDMA (time division multiple access) scheme is used, through which the OLT gives permission to only one ONT to transmit upstream data at a given time.

As an example for the rates supported by the current PON standards, given the above considerations, a 2.5 Gbps GPON access implementation servicing 32 customers can support about 78 Mbps per customer (if divided equally), which drops to about 39 Mbps per customer if the total number of customers is doubled to 64. Most current commercial GPON systems support an operating distance of 20Km if a split ratio of 1:32 is used, which drops to 10Km if the split ratio is increased to 1:64.

As can be perceived, current PON standards can support higher and guaranteed bandwidth to customers and much larger operating distances than xDSL networks. However, their main drawback lies in their inherent operation, and specifically to the fact that they rely on simple optical power division amongst different customers serviced from the same feeder fiber and the associated necessary use of TDM methods. This limits the maximum achievable bandwidth per customer, the number of customers that can be serviced within a PON branch, and the operating distance of the network. In addition, the use of TDMA mechanisms for the time-scheduling of upstream traffic limits the available bandwidth per customer in that direction, and adds to the overall complexity of the system.

It is generally accepted that for the future upgrade of PON implementations capacities, as well as an increase of the supported end-connections, CWDM or even DWDM technologies will be employed, hence leading to flexible WDM-PONs featuring very large bandwidth .

3 WDM-PON Technologies and architectures

With the continuously increasing need for efficient provisioning of bandwidth-demanding new services, it is almost certain that the current PON access networks will evolve to WDM-PONs featuring increased number of wavelengths for larger capacity (Kani et. al., 2003). In the ultimate vision of a WDM-PON, every ONT will be able to communicate with the OLT via its own dedicated wavelength pair, thus enjoying the full fiber's bandwidth of a single wavelength (corresponding to a point-to-point (P2P) link) and bypassing the inherent limitations of TMD-PONs discussed above. The incorporation of increased WDM characteristics into PON technologies, apart from the obvious bandwidth increase, also allows for a rich enhancement of the network configuration and dimensioning properties (e.g. service differentiation on a wavelength basis, layering of different PON networks etc). It is noted that WDM-PONs are not standardized yet and are currently intensely being investigated by the academic and industrial R&D communities.

Figure 1 displays a simple illustration of a generic WDM-PON featuring one-stage splitting. As can be seen, and in contrast to the current PON standards, each ONT is serviced through a dedicated wavelength pair. The optical power splitters of current TDM-PONs have been replaced by "smarter" (though still passive) wavelength multiplexers/demultiplexers (e.g. grating structures) that route the traffic according to wavelength. In the downstream direction the OLT sends traffic of multiple channels of different wavelengths that are separated by the WDM-splitter and routed to the appropriate ONTs. In the opposite direction, every ONT sends traffic using its dedicated upstream wavelength. All upstream channels are multiplexed together by the WDM-splitters and sent to the OLT via the common fiber link. Overall, as can be perceived, although a common infrastructure is largely used, the link between the OLT and each ONT is essentially P2P, with no bandwidth division between the supported users and no need for traffic time-scheduling mechanisms. Traffic scrambling may also be not necessary.

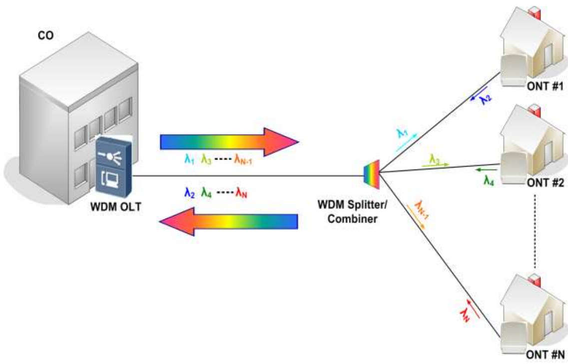


Figure 1: Schematic of WDM-PON operation in which each ONT uses a dedicated wavelength pair for communication with the OLT.

The maximum capacities that can be delivered to end-customers with such schemes are truly astonishing (of the order of many Gbps) with respect to the demands of current access networks. However, it is noted that for such schemes and in order to support many users per PON branch, the employment of DWDM technologies is necessary, which prohibitively rises the implementation costs, at least with today's standards. In addition a number of major technical complications arises (notably such as the need for "colourless" ONTs at the user side) (Davey et. al., 2006). Hence, currently the employment of CWDM technologies seems much more favourable in order to start the implementation of the first WDM-PONs, which however pose strict limitations on the network dimensioning and operation due to the limited number of available wavelengths.

There are currently several approaches with regard to the introduction of CWDM technologies in PON networks. The greatest interest in terms of functionality supported and future commercial exploitation is presented by proposals in which the provided services are differentiated on the basis of wavelength, or in which operation of multiple overlaid PON networks is achieved through a common infrastructure. The first case is illustrated in Figure 2(a), in which the various provided services (e.g. fast internet, IPTV etc) are assigned a different operating wavelength. It is also possible to assign multiple wavelengths to particularly demanding services if such a need arises. Excluding the need for active CWDM equipment in the OLT and the respective separation filters in the ONTs, such networks (more precisely their passive infrastructure) are otherwise identical with the current TDM-PONs of optical power division. The use of additional wavelength greatly increases the available capacity, and allows to meet the bandwidth and QoS requirements of demanding services (such as for example multi-channel IPHDTV). Also, new services can be introduced in the network rapidly and efficiently. In the second case (overlaid PONs, see figure 2(b)), CWDM technologies can be used to combine several PON networks in a unified physical topology. In such implementations, there is an initial division of a multitude of wavelengths that correspond to different PONs, and afterwards an intra-network optical power division for the traffic to reach the end-users. In this way, a large increase of the maximum number of supported users per feeder fiber can be achieved. This, however, still requires the use of “coloured” lasers (i.e. lasers of different wavelength) at the ONTs of each overlaid PON branch, a fact that increases the cost and complexity of the implementation.

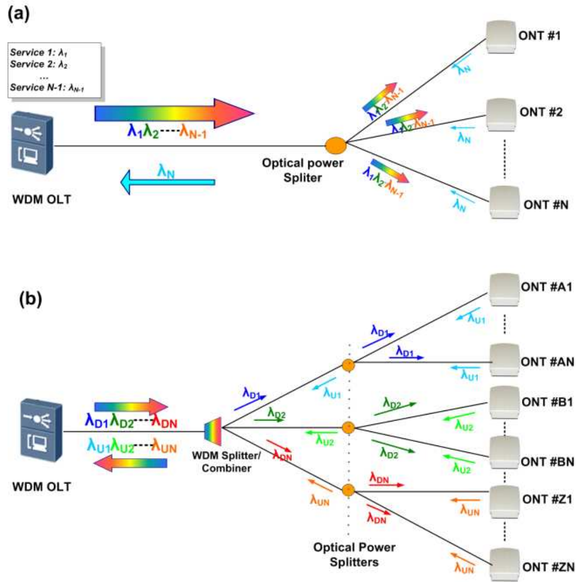


Figure 2: Architectures showing the implementation of CWDM technologies for (a) service provisioning with wavelength differentiation, and (b) creation of overlaid PON networks over the same infrastructures with a dedicated wavelength pair per PON.

A slightly different commercially interesting variation of the above case would be to use CWDM technologies to combine several existing GPON or EPON networks, as well as the creation of hybrid TDM/WDM PONs (An et. al., 2005).

4 WDM-PONs For wireless backhauling

A WDM-PON seems to be an excellent technology for managing, distributing and backhauling high bandwidth data coming from different wireless access points. In the case that the wireless station is based on LTE or WiMAX this can generate a large amount of traffic that when it is aggregated with traffic coming other wireless points, then the huge amount of traffic can be transported in a rather flexible way by using WDM PON structures as the one seen in figure 3.

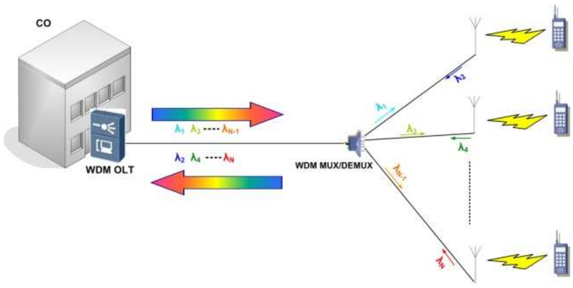


Figure 3: A WDM PON architecture for transporting high pit rate wireless services

The above combined treatment of optical and wireless media results to enhanced flexibility and granularity options towards ensuring bandwidth and QoS guarantees. The WDM – PON architecture can offer Superframe structures of constant duration for every wireless station (WS) and one of constant number of DataFrame-within-Superframe allocation per user for maximum bit rate usage. The constant Superframe duration will enable guaranteed bandwidth at every WS offering increased throughput values where less end-users are served, whereas the constant number of DataFrame allocation per user can provide optimized bandwidth sharing between the WSs at a user-dependent level, allocating more bandwidth at WSs with increased number of served users whilst reducing the allocated capacity at WSs with less users. This MAC scheme of the above architectures will aim at optimized Superframe structuring and distribution algorithms with respect to the type of traffic requested (wired or wireline) taking advantage of the centralized network management carried out at the CU where in order to exploit the maximum possible feedback information about the network's load-per-service conditions.

An obvious advantage that comes along with the huge bandwidth offered is the high capacity for the RF signal transmission. This capacity between the RAU and the CU is able to satisfy new wireless broadband technologies such as Advanced-LTE and mobile WiMAX that promise performance rates superior to those of 1 Gbps. Furthermore, the available bandwidth that is in excess of that achieved by using copper cable or broadband wireless systems gives the opportunity to carry voice signals, simple data and multimedia (video on demand, IPTV, HDTV, online gaming etc.) traffic eliminating the backhaul bottleneck problem.

The bi-directional rapid information transfer from the Central Unit to wireless users and Base Stations and vice versa, enables the accurate coordination of network architectural elements to achieve the desirable outcome. The WDM-PON architecture comprising only of passive optical components that do not require power supply, eliminates major installation and operating cost.

Acknowledgement

The authors would like to acknowledge the partial support of the European Community in the framework of FUTON FP7 ICT-2007-215533 project for writing this paper.

REFERENCES

An, F-T., Gutierrez, D., Kim, K.S., Lee, J.W., Kazovsky, L.G., 2005. SUCCESS-HPON: A next-generation optical access architecture for smooth migration from TDM-PON to WDM-PON, in *IEEE Communications Magazine, vol. 43, no. 11, pp. 40–47*

Davey, R., Kani, J., Bourgart, F., McCammin, K., 2006. Options for Future Optical Access Networks, in *IEEE Communications Magazine, vol. 44, 10. 2, pp. 50–56*

Institute of Electrical and Electronic Engineers (IEEE), 2004. *IEEE 802.3ah Standard Ethernet in the First Mile*, New York: IEEE.

International Telecommunication Union - Telecommunications Standardization Sector (ITU-T), 2001. *ITU-T Recommendation G.983.1: Broadband optical access systems based on Passive Optical Networks (PON)*, Geneva, Switzerland: ITU-T.

International Telecommunication Union - Telecommunications Standardization Sector (ITU-T), 2003. *ITU-T Recommendation G.984.1: Gigabit-capable Passive Optical Networks (GPON): General characteristics*, Geneva, Switzerland: ITU-T.

Kani, J., Teshima, M., Akimoto, K., Takachio, N., Suzuki, H., Iwatsuki, K., and Ishii, M., 2003. A WDM based Optical Access Network for Wide-Area Gigabit Access Services*, in IEEE Communications Magazine, vol. 41, no. 2, pp. 43–48.*

ADDED-VALUE SERVICES PROVISIONING OVER INTER-TECHNOLOGY RADIO ACCESS NETWORKS

M.Tsagkaropoulos[1*], I. Politis[1], C. Tselios[1], T. Dagiuklas[2] and S. Kotsopoulos[1]

1. Dept. Of Electrical and Computer Engineering, University of Patras, Patras, Greece
2. Dept. of Telecommunications and Network Systems,Technical Institute of Messologhi, Nafpaktos, Greece
{mtsagaro, ipolitis, tselios, kotsop}@ece.upatras.gr, ntan@teimes.gr

Keywords: IMS, NGN, MIH, RATs PAIL, advertisement selection, CSA, VSM

Abstract: Today's mobile communication sector philosophy is the Always Best Connected (ABC) philosophy, which is based on the fact that several co-existing Radio Access Technologies (RATs) with different requirements and capabilities are 'competing' for best candidate network. For an advanced scenario these RATs can range from existing technologies (e.g. UMTS, WiMAX) to future networks. Smooth inter-working among heterogeneous wireless technologies can be accomplished by deploying the Media Independent Handoff (MIH) 802.21 framework. In addition, the convergence of heterogeneous wireless technologies over an All-IP core network is allowing mobile subscribers to access a variety of services, access networks and devices. The IP Multimedia Subsystem (IMS) is the unified control framework that will support multimedia services and provide seamless service delivery, availability and continuity across different domains and access technologies.

This study presents the deployment of value-added services over interworking heterogeneous wireless networks. A new entity called Personalization and Advertisement Insertion Logic (PAIL) is introduced, which enables a mobile network operator to exploit contextual data stored in its network for personalized advertisement selection. The convergence and interoperability of heterogeneous mobile and broadband network technologies are discussed along with proposed solution framework for targeted advertisement in IMS. An evaluation of the advertisement selection is presented which includes user tests.

1. INTRODUCTION

Today's deployment of Next-Generation Networks throughout the world enables the provision of multimedia applications with higher requirements in terms of quality of service and connectivity. To satisfy these demands a variety of access technologies has become available: WiFi, UMTS, WiMAX etc. Thus, service providers are adopting the Always Best Connected (ABC) philosophy (Dahlman E., 2007), which is based on the fact that several co-existing Radio Access Technologies (RATs) with different requirements and capabilities are "competing" for best candidate access network. It is then necessary to have mechanism that, for a particular instant in time, select the "best" network and perform the inter-working between different technologies. Smooth inter-working among heterogeneous wireless technologies can be accomplished by deploying the Media Independent Handoff (MIH) 802.21 framework (IEEE 802.21, 2008).

In addition, the convergence of heterogeneous wireless technologies over a broadband IP core network is allowing mobile subscribers to access a variety of services, over a variety of access networks and by using a variety of devices. Hence, there is the need to consider IP Multimedia Subsystem (IMS) as the unified control framework that will support multimedia services and provide seamless service delivery, availability and continuity across different domains and access technologies.

The major disadvantages of existing mobility solutions are, that they consider link-layer and network layer triggers in order to carry-out handover. They do not capture the application-related QoS requirements in order to support session continuity and maintain QoS of on-going sessions.

In the case of mobile advertisement, it is important to provide targeted advertisement whenever a user is switching between RATs. Each access network has certain limitations that effect users' perceived quality in terms of throughput, delay, jitter etc. As a consequence the provided services may be limited or degraded. Hence, it is important to include a mechanism that enables the service provider to advertise services and applications that are supported by the connected RAT whenever a user is getting connected.

This paper presents MIH as middleware mediator, which includes the handover selection mechanism on top of an intergraded MIH and IMS system. The MIH collects statistics from different layers that are imported in the decision handover algorithm and determine where on going session should be handed off by taking into account various cost parameters. Moreover, this study focuses on the Advertisement Selection Scheme used in the IMS-MIH integrated environment. The scheme uses a mathematical model in order to perform the user profile and advertisement metadata correlation so as to select the appropriate personalised advertisement when the user handoffs to a new RAT.

2. EVOLUTION IN NEXT GENARATION NETWORK CHARACTERISTICS

In the context of the ubiquitous seamless service mobility, which is responsible for interoperability and cooperative control providing services, it is important to define the framework (in terms of mobility protocols), allowing a MT to handover uniformly among heterogeneous link-layer interfaces. This infrastructure will support the dynamic establishment of trust relations between independent providers (e.g., foreign and home providers) in a distributed manner over hybrid IPv4 and IPv6 networks (Salkintzis A., 2004).

Moreover, it will provide the required enhancements for providing secure interconnection among different heterogeneous networks, establishing user-provider trust relations and the necessary means for authenticating users in foreign domains and exchanging their profiles in a secure manner. This would thereby enable users to roam to foreign networks and use the provided services in these networks without affecting their privacy.

Such challenge can be met by Media Independent Handover (MIH) mechanism: optimistic seamless mobility and maintain session continuity. IEEE Working Group has recently proposed IEEE 802.21 standard to enable handover and interoperability between heterogeneous networks with context-awareness in mobile terminals (G. Lampropoulos, 2008). One of the main ideas behind IEEE 802.21 is to provide a common interface for managing events and control messages exchanged between networks devices that support multiple interfaces, both wired and wireless. MIH considers a co-operative decision making framework by taking measurements and triggers from both mobile terminal and radio access network (Tsagkaropoulos M. et al, 2009).

On the other hand, the IP Multimedia Subsystem selection (3GPP TS 23.228, 2007) has been developed in order to provide a service delivery platform (SDP) for a converged communication paradigm. The target was the convergence of voice and data networks under a single communication platform on top of different network access technologies. The IMS standards focus on the network aspects of service delivery, including both network control (signalling traffic) and network delivery (content traffic) of multimedia services.

The key components of IMS are distributed among service planes that categorize the functions provided, including user, access, network, transport, control and service/application planes. Devices in the user plane establish a SIP-encapsulated IP connection to applications situated within the service plane, supporting a variety of access networks (e.g. GSM, wireless area networks, etc.). More details on IMS can be found at (Camarillo G, 2008).

The IMS standards include the roaming ability where the visited network operator has a roaming agreement with the home network operator. In these agreements both operators negotiate some aspects of the service provided to the user, such as the price of calls, the quality of service, or how to exchange accounting records. 3GPP is currently designing a way to implement handover with the so-called Voice Call Continuity (VCC) within the concept of Fixed-Mobile Convergence (FMC) (3GPP TS 23.206, 2007). Our approach tries to merge current IMS deployment with the upcoming IEEE 802.21 framework, with minimum changes on the system architecture and within the concept of centralized services for all clients. In order to extend the functionality of IMS across different access domain with focus on seamless mobility and service continuity, the MIH functionality is used as a middleware mechanism that interconnects IMS domains in case of handoff. An overview of the system is illustrated in Figure 1.

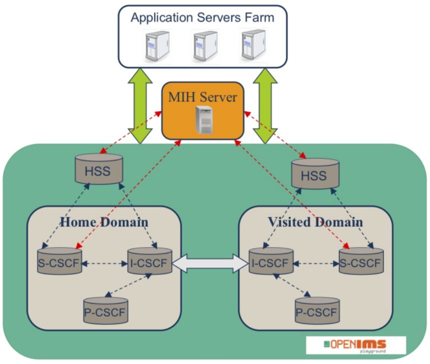


Figure 1: Interconnection of MIH and IMS

3. MEDIA INDEPENDENT HANDOVER IN IMS

The IEEE 802.21 standard aims to enable handovers between heterogeneous technologies including WLAN, 3G and other wireless access technologies without service interruption, hence improving the user experience. Contrary to the current functionalities that provide service continuity with the cost of complex interactions specific to each particular technology, IEEE 802.21 provides a framework that a ensures simplified interactions among higher and lower layers for achieving session continuity independently of the underlying technologies.

In particular, the 802.21 framework primary intention is to enable seamless handoff and interoperability between heterogeneous network types. This is done by introducing the layer 2.5 specified by Media Independent Handoff Function (MIHF) which provides three principal functionalities: Media Independent Events Service (MIES), Media Independent Command Service (MICS) and Media Independent Information Service (MIIS).The interactions among lower and upper layers and the role of the MIH functions are illustrated in Figure 2.

As it is explained, the 802.21 framework is deployed in the MIH central unit which is responsible for monitoring different RATs and controlling network discovery, resource management, link configuration and selection. The key module of the MIH functionality is the MIH server which hosts the Handoff functionality (HoF) and decides upon the necessity of a handoff based on Application and higher layers parameters. Moreover, the MIH server interacts with the Mobile Terminal (MT), requesting application, network and transport layer parameters that are important for the handoff decision algorithm. Finally, the MIH server interacts with the Mobile IP and the Link Selection modules, informing both modules that a handover has been decided.

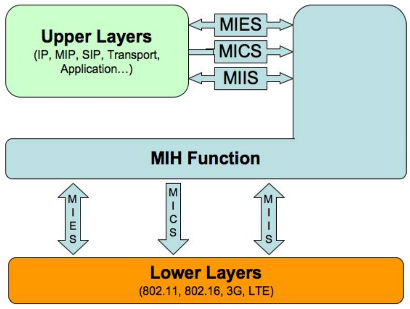


Figure 2: Layer interactions via MIH key functions

The Handover functionality (HoF) is part of the MIH server that is located in the middleware. The handover algorithm itself utilizes parameters that are collected in the HSS database and consists of the following stages:

– Handover decision: The MT related information (e.g. SNR, delay, jitter, PSNR, packet loss, etc), are collected and sent to the MIH server via the MIH information service. In addition, the MIH information service is responsible of collecting corresponding information from neighbouring networks, in order to be used during the handover decision-making.
– Handover initiation: These collected MIH parameters will be evaluated and compared against a set of predetermined threshold values. These thresholds are either determined be the network provider and are specified in each user's profile. The MIH event service is responsible for comparing the collected statistics with the threshold values and for informing the command service when one or more thresholds are violated and the handover criteria are matched. The Link Selection module indicates the best candidate neighbouring network for handoff according to specified target QoS parameters. The target QoS parameters are specified by the service provider according to the Quality of Service classification of the user (gold, silver, bronze).
– Handover execution: The final stage is the execution of the vertical handover to the decided neighbouring network.

In our interworking architecture, the MIH server is the trusted entity between the two IMS domains, the home domain and the visited domain. The MIH server is responsible to check the status of each MT as concerns as link performance and quality of service. When the handover is initiated the MIH server is responsible to initiate a RE-REGISTER message that will transfer the current session from the home domain to the visited domain.

Since the MT is accepted at the visited domain, then the MIH server triggers the MT to update his network specifications so as to forward his traffic through the visited domain. Finally, a SIP UPDATE message is forwarded to MT requesting to update his session according to the new domain specification and a RE-INVITE message is forwarded to the new S-CSCF in order to continue the session.

When the MT is accepted in the visited domain, he is able to receive advertising information for the services and applications that are appropriate for him based on the following framework

4. FRAMEWORK FOR MOBILE ADVERTISEMENT

4.1 Limitations of modern mobile advertisement

The introduction of mobile advertisement includes two aspects that might be considered of great importance. When user information for example location is accessed by advertisers, questions regarding privacy rising. In addition such user oriented functions might have an impact on network scalability or even its robustness. Therefore it is essential for this approach to be fully evaluated and justified before actually implemented in large-scale network simulators.

Behavioural targeting advertisement techniques often neglect a rather important user right, the right of privacy. Companies offering advertisement services, by using monitoring methods might cause people to feel uncomfortable, provided they are aware of what exactly is takes place while they are using prepaid or not telecommunication services. User data is scanned with great detail extracting results which are taken into consideration for discovering possible patterns revealing human behaviour and socialization schemes. The problem of personal data misuse is gaining attention, also by legislative bodies. But still consumers can never be certain what might happen with their data, once they are given away.

The relationship between users and telecommunication carriers is different than the one mentioned above, in the sense that a legal bond in contract form exists between both parties that regulates the use of personal data. The information that Carriers get could be given to the advertisement companies is provided in full anonymity and could still be used for the benefit of the user - for example by introducing certain business models that lowers usage costs in exchange for the possibility of personalized advertisements.

4.2 IMS featuring PAIL

4.2.1 The IMS framework

Being the unified control framework that supports multimedia services and provides seamless service delivery, availability and continuity across different domains and access technologies, the IP Multimedia Subsystem (IMS) is considered to be the ideal playground for deploying a new entity, which enables a network operator to exploit contextual data for personalized advertisement.

In the case of IMS user data is stored in the HSS database operated by the network provider. All data regarding a single user including device characteristics as well as presence is available to the provider and it only takes a module having universal access to the specific repositories of each domain to acquire all kinds of information and use it for service enhancement. The IMS-MIH integrated environment allows significant data flow within different domains thus making the whole effort even easier.

4.2.2 PAIL in the IMS domain

The new entity introduced is called Personalization Advertisement Insertion Logic (PAIL) and aims to enhance different existing services, live and non-live (e.g. Podcast, Mobile TV). Its task is selecting suitable advertisement material which is then inserted by other entities. The PAIL design itself is platform independent and the actual integration with the IMS framework is shown in Figure 3

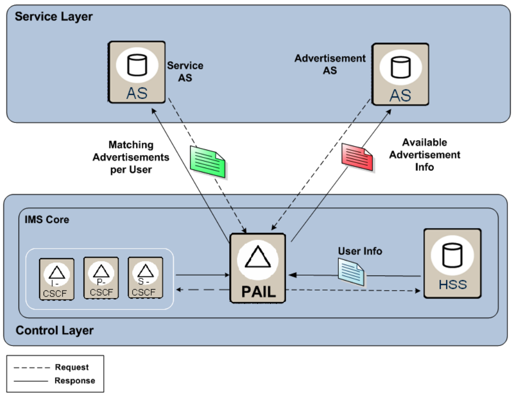


Figure 3: PAIL in the IMS

The aforementioned architecture spreads across two out of the three Layers that exist in the IMS Architecture: the Control Layer, where all modules operated by the network provider are located, and the Service Layer that contains all kinds of Application Servers (AS), some created by the network provider in order to enable fundamental services and others by 3rd party service providers. PAIL communicates with the necessary entities for getting user information and it is also connected to two distinct AS of the Service Layer via specific interfaces. The first application server, Service AS, represents an end user service like e.g. podcast while the second, Advertisement AS, is an advertisement clip metadata repository, where information for available advertisement clips is stored. The other two entities located in the IMS domain are HSS and the three types CSCFs.

The internal PAIL architecture is shown in

Figure 4.

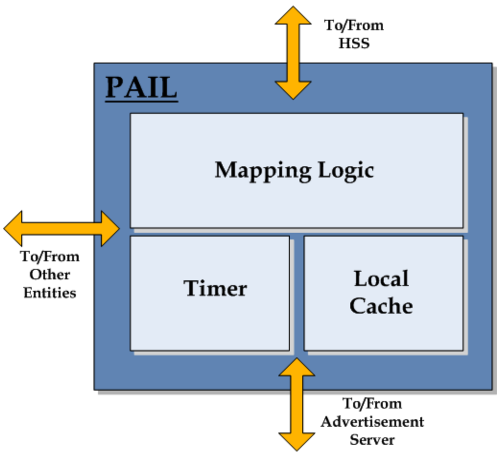


Figure 4: PAIL Architecture

Inside PAIL takes place the correlation of information regarding advertisement clip metadata and user profile, in order for the best possible match of advertisement per user to be chosen. The module remains in standby (or idle) status until a request for personalized advertisement arrives from an AS. After issuing requests for data towards all other necessary entities, PAIL sets up a Timer and switches to “Wait for Response” mode. In addition, it uses a local cache for temporary storage of advertisement meta descriptions and user profiles it receives. Communication to the network entities, CSCFs and HSS is done via the standardized interfaces these modules use as described in 3GPP TS 29.228 (2004)

4.2.3 Content Selection Algorithm

The Content Selection Algorithm (CSA) located in the Mapping Logic function is responsible for the actual matching process between profiles and advertisement material that makes personalized content delivery possible. It is based on a mathematical model called Vector Space Model (VSM) in order to perform the previously mentioned correlation. VSM is an algebraic model for representing text documents as vectors or identifiers (Grossman & Frieder, 2001). Documents are represented in such a way that in order to compare their similarity, all it takes is to check the corresponding vectors.

Since all documents consist of words, numbers and a finite variety of symbols there is a method for all components to be mathematically expressed. It is possible to create a list of every single word that exists in a document collection and represent all documents as series of numbers that symbolize the total amount of times a specific term can be found in the document. After the series of numbers is created, it can be treated as a vector, applying all rules of vector analysis on it. Several different ways of computing these values, also known as weights, have been developed (Lee, Chuank, & Seamons, 1997).

In a true case scenario, the total vector space dimension is equal to the number of different terms existing in the whole document collection. A user, whose profile has many identical words with an advertisement description, is more likely to be interested in this advertisement content. This is exactly the clip that CSA is going to match to the User ID. Based on the nature of this scenario, which involves free text description within user presence data, the best VSM approach out of many that have already been implemented in the past (Lee, Chuank,

& Seamons, 1997), was considered to be the full Vector Space Model.

An advanced attribute of the CSA over the simple VSM is normalization. The number of times a single term can be found in a document is divided by the total number of document terms. In addition, weight normalization (Lee, Chuank, & Seamons, 1997) is used, where a specific term is divided by the total number of times a term exists in the whole document collection. This method provides more accurate results, and contains benefits when vectors tend to increase in size and common terms are rendered unimportant when compared to unique or rare words that might even define a specific document. For instance, terms like articles or verbs being of minor importance for the document definition compared to nouns or adjectives, get lower weights after normalization.

In conclusion, normal VSM is obviously capable of performing simple correlation tasks but for optimal performance and better results the CSA approach prevails.

5. EVALUATION SCENARIOS

In order to properly evaluate the influence of additional data provided by the operator, emphasizing on location as the most important addition to the user profile, two discrete test scenarios are introduced. In the first scenario under the name Operator's Data Approach, input data regarding the user consists of his/her public profile and carrier data (here: location), while in the second scenario the user profile alone was used. The goal is to evaluate if more comprehensive information actually leads to a better match of selected advertisements and user interest. The CSA is based on an implementation of an information retrieval system.

For the moment there is no automated procedure for acquiring user profiles; therefore they must be manually inserted as an input. To measure user satisfaction with the proposed algorithm in a realistic setup, a video podcast playback along with the proposed advertisement clips that the algorithm extracted, was used. For this task five podcast categories were chosen: Sports, News, TV Series, Comedy and Music Video. These categories were selected to span a broad range of themes that most users are interested in. By picking rather general podcasting categories, the influence that this content might have on the subject's impression of the advertisement clips was minimized.

In addition, ten advertisement clips were chosen separated into two groups, but this time different clips for each user according to the correlation of user profiles to ad clip metadata performed by the CSA. A group of eighteen test volunteers was selected and each one provided a description of his/her personal preferences as part of the user profiles. It should be pointed out that no subject is aware of the actual scenario that he/she currently experiences.

A point system together with actual clip satisfaction questions was used to compare the two scenarios. In those questions given to the participants, a scale of five values was offered as answers. The higher the collected point number the more satisfactory the experience was meaning that the advertisement was somehow closer to his/her preferences. The clip questions were:

1. Was the advertisement clip interesting for you?
2. Would you like to get more information about the product?
3. Would you be willing to buy the product if such chance was given?
4. Was the advertisement clip well-produced?

As shown in Figure 5 the actual point assignment of each user per question clearly indicates the participants' satisfaction in the first set of podcasts where additional data originated from the carrier were user for the advertisement selection to be realized.

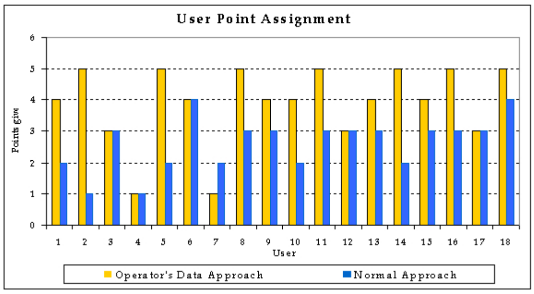


Figure 5: User Point Assignment

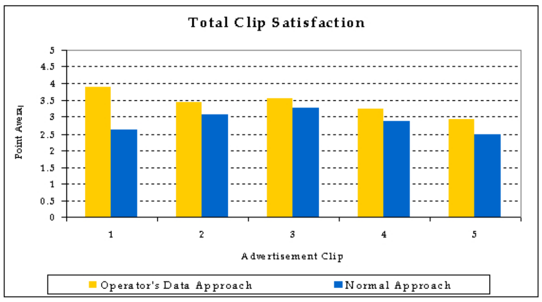


Figure 6: Total Clip Satisfaction

By accumulating the average points given to each clip by users some interesting results regarding total clip satisfaction were extracted, as shown in Figure 6. In all cases users selected to assign more points expressing satisfaction to the advertisement clips chosen by the CSA. This strengthens the original hypothesis of the extra value of information provided by the network carrier in order to deliver a much better service to the end user. Indeed all users seem to enjoy the first session better than the second one since the selected advertisements considered to be more interesting and closer to their preferences.

6. CONCLUSIONS

The introduction of PAIL in a next generation network framework is about to make operators capable of providing better services and revolutionize the overall user experience. As part of a wider architecture currently under development that will allow real-time decisions to be taken, this module is the initial step for a far more complex solution that will not only correlate users and advertisement but also use all available network information in terms of resources for making decisions on how content delivery is going to be achieved. Acting in control of delivery modes such as broadcast and unicast and also being aware of possible user groups that might exist according to preferences and content, it will be capable of executing optimal delivery services as well as advertisement selection

ACKNOWLEDGEMENTS

This work was partially supported by the European Commission under the Grant Agreement No. 21533 (*FUTON: Fiber Optic Networks for Distributed, Extendible Heterogeneous Radio Architectures and Service Provisioning*), FP7 ICT-1-1-1 Activities (Network of the Future).

REFERENCES

3GPP TS 23.206 (2007). *Voice Call Continuity (VCC) between Circuit Switched (CS) and IP Multimedia Subsystem (IMS).*

3GPP TS 23.228 v7.9.0; (2007) *IP Multimedia Subsystem (IMS); Stage 2 (Release 7).*

3GPP TS 29.228 v6.5.0. (2004) *IP Multimedia (IM) Subsystem Cx and Dx interfaces; Signalling flows and message contents.*

Camarillo G. and Garcia-Martin M (2008). *The 3G IP Multimedia Subsystem (IMS): Merging the Internet and the Cellular Worlds.* 3rd ed. John Wiley & Sons Ltd.

Dahlman E. et al. (2007). *3G Evolution: HSPA and LTE for Mobile Broadband.* Elsevier.

G. Lampropoulos, et al (2008). Media-Independent Handover for Seamless Service Provision in Heterogeneous Networks. *IEEE Communications Magazine.* Vol. 46. no. 1. pp. 64-71. January.

Grossman D., & Frieder O. (2001) *Information Retrieval: Algorithms and Heuristics.* Second Edition, Springer

IEEE 802.21/D10.0.(2008) Draft Standard for Local and Metropolitan Area Networks: Media Independent Handover Services. *IEEE Draft.* April.

Lee D. L., Chuank H., & Seamons K. (1997) Document Ranking and the Vector-Space Model. *IEEE Software*, 0740-7459

Salkintzis A. (2004). Interworking Techniques and Architectures for WLAN/3G Integration towards 4G Mobile Data Networks. *IEEE Wireless Communications*, June, pp. 50 - 61.

Salton G. & Buckley C. (1990) Improving Retrieval Performance by Relevance Feedback. *Journal of the American Society for Information Science*, 41(4), 288- 297.

Tsagkaropoulos M. et al (2009), *Evolution of Next Generation Networking for secure multimedia services: The MIH approach*, 48th FITCE Conference, Prague, Czech Republic, 3–5 Sept.

OPEN SOURCE SOFTWARE AND ONLINE COMMUNITIES INTERACTION IN THE FINNISH INDUSTRY

Preparation Contributions to FITCE 2010 Proceedings

Víctor Acinas Garzón

Faculty of Business and Technology, Tampere University of Technology, P.O. Box 541, FI-33101 Tampere, Finland

victor.acinasgarzon@tut.fi

Keywords: open source software, customer perception, online communities, software component, revenue logic

Abstract: The software industry is recently undergoing deep changes linked to open source software (OSS), online communities and customer perception. Companies can no longer rely on in-house software development as a competitive advantage and should look towards open source software (OSS) and online communities as a competitive software development methodology. The paper is based on an interview study research approach that consists of an extensive literature review in the field and theme interviews with OSS industry specialists. Because of this, the paper shows industry trends related to the topics considered relevant in the existing literature, such as the non-differentiating or differentiating software component focus, and detect new and/or emerging topics. Hence, the findings show whether companies should focus on non-differeatating or differentiating software components. As well as the importance of harnessing online community developments to align the company's business model and tap the dyadic relationship among OSS and proprietary software.

1 INTRODUCTION

Despite the fact that software is taken for granted in our everyday life, software dates back to approximately only half a century. Interestingly, at that time hardware vendors distributed software for free with their products and the only possible business was to offer tailored services to the professional segment. Later, hardware and software decoupled enabling the appearance of the enterprise solutions market such as ERP and CRM. The next major market was packaged software, triggered by the development of the personal computer (Hoch et al, 2000). As a result of the numerous personal computers and Internet, customers arrange themselves in online communities (McDermott, 2001).

New development methodologies for software challenge the traditional methods (Chan & Lee, 2004). An example of these new development methodologies is open-source software (OSS), which, in contrast to proprietary software, provides the source code used to customers. Thus, the current software market offering can be categorized according to code access (open or closed) and price (free or non-free), see Figure 1.

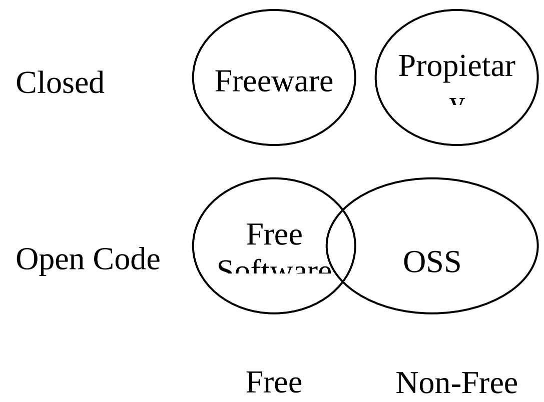


Figure 1. Software categories attending to price and code accessibility.

Hence, OSS opens up new possibilities as it gives the opportunity for *anyone* to evaluate, modify, and tailor the software according to his/her needs. Thus, the user is allowed to re-distribute or, for instance, sell the software resulted from the tailoring as a component of a package (Coar, 2006).

However, companies find it difficult to implement the changes OSS and online communities involved in their organizations, mainly because the development process and the resulting outcomes of these (new) development methodologies do not follow traditional software development processes and structures.

2 Online Communities

The information exchange increase among consumers and businesses can no longer be controlled in a traditional manner, one-way, as they are leveraged through on-line services. Nowadays, more and more consumers are able to communicate among each other leading to i.e. discussions, forums, wikis, and, ultimately, online communities. Hence, businesses can arrange consumers in communities to cope with the information flow. Community is a…

…group of people with a common purpose whose interaction is mediated and supported by technology and governed by formal and informal policies (Preece, 2000).

In addtition, communities can be generated spontaneously by users or strategically by companies (McDermott, 2001). Strategic communities' interaction and participation bridges the gap between consumers and companies (Ramirez, 1999; Prahalad & Ramaswamy, 2002). For instance, the Open Directory Project community managed and designed by Netscape generates, for example, innovation, technical knowledge, and business value (Pavitt, 2002 and Brusoni et al., 2001). On the other hand, a well-known example of a spontaneous community is Linux that exhibits a fast development and exchange of knowledge among the participants.

Hence, companies seek different advantages when planning the creation of an online community. The main objectives behind an online community can be summarized as follows:

- Improve the product development process with special emphasis in idea generation, product conceptualization and prototyping (Chan & Lee, 2004).
- Enhance innovation in the creation of new products and designs to enhance the product characteristics and specifications (Chan & Lee, 2004).
- Provide ways of offering product testing and support other software users (Cusumano, 2004).

In all, online communties have become a key aspect of the company's operation model of software development as it intertwines with the OSS development. This new development methodology proposes a win/win situation for the community and the business. It is in the best interest of both to sustain this relation, in contrast to traditional development processes "planning, composing and testing" (Chan & Lee, 2004) that only rely on in-house resources.

3 OSS REVENUE LOGIC AND STRATEGY

Companies often approach OSS revenue logic models with a cautious attitude despite the growth of OSS due to the singular involved characteristics such as free distribution of the source code and modificability. Furthermore, OSS revenue logic strays from traditional software in various ways and does not rely on conventional proprietary software licensing models. The identified revenue logic models are described in Table 6.

Table 6. OSS revenue logic models (modified from Hecker, 1999 and Välimäki, 2005).

Revenue Logic	Description
Support	Software is distributed free of charge and a company is responsible for the support
Commodification	The product evolution in the technology life cycle (see Figure2) leads to OSS.
Brand licensing	Though the initial product is OSS any derived products that use their brand names or trademarks must compensate the creator.
Loss-leader	Companies may introduce non-profitable proprietary products as to gather attention, demand, improve quality... while expecting other sources of revenue.
Services	All those services derived from the usage of OSS such as consulting or on-line services.
Accessorizing	Company's that directly or indirectly benefit from offering supportive OSS such as books, hardware and textiles.

The previous revenue logic models serve as a framework to provide guidance in conjunction with the direction that OSS development shall follow. In addition, since software and the environment are constantly changing, companies may find suitable to change their revenue logic model whenever it is convenient. For instance, at first a *Loss-leader* revenue logic may be followed but, as the gathered attention and interest soar, a company may switch to others such as accessorizing, support or services. Furthermore, as a critic to Hecker (1999), the *Loss-leader* and *Commodification* strategies, do not follow the definition of revenue logic. Thus, those strategies shall be considered as a transition or the means toward other ends.

The strategic implications of OSS as a new developing methodology are first, opening the source code could make companies loose the differentiating attributes of their products/services and, second, allocating resources on software component that provide no added value and result in a waste of resources, commodity.

A commodity refers to basic components that can be used for different purposes within software. Interestingly, a commodity may have been regarded as novel and innovative component when introduced. For this reason, products such as cars, personal computers and portable music players are regarded as commodities while once were differentiating.

Despite the short life of the software industry some components have passed from being considered as a competitive advantage to a cost of doing business as they move along the technology cycle. This evolution is a result of the competitive pressure of the business to take advantage of new components and no longer being regarded as a premium product or service but, on the contrary, considered as a commodity. To name a few, Van der Linden et al. (2009) and Stutz (2004) refer to this trend as the commodification of software.

Figure 2 illustrates how software's technology life cycle evolves as a consequence of commodification. Thus, technology categorization changes (that is, the move from top to bottom along the vertical axis) in conjunction with the most efficient and effective schemes of collaboration (that is, the move from left to right along the horizontal axis).

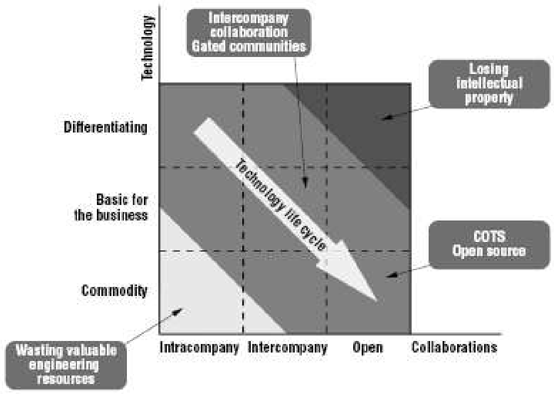


Figure 2. Efficient and effective software development (Van der Linden et al., 2009).

In addition, Figure 2 underpins the importance for companies to understand and manage not only OSS development methodologies but also Commercial Off-the-Shelf Software (COTS) and the communities involved to increase their efficiency and effectiveness of the company. According to Chan & Lee (2004), OSS online communities can achieve differentiating software components against Van der Linden (2009), who regards it exclusive of in-house intensive developments.

4 Research method

This paper study is based on the interview research methodology. Thus, the research process starts with a thorough study of the existing literature and relevant business cases. Interestingly, most of the online communities for OSS development discussed by Chan & Lee (2004) have used software component as a differentiating factor even though Van der Linden (2009) emphasizes OSS should mainly be used for development of non-differentiating software.

This provides an interesting starting point for interviewing the OSS and software experts in Finland in order to understand and analyse the opinions of industry specialists on this matter through theme interview (Carson et al, 2001; Merton et al., 1956). In this manner, trends in the Finnish industry are identified in relation to the topics considered relevant in the existing literature and emerging topics in the field can be detected.

5 Analysis

In this section, the paper analyses the main results of the study. The following subsections identify the main results along with the comments and company cases that the interviewees pinpointed.

5.1 New Ways of Utilizing OSS and Online Communities

The interviews indicate a fast growth of OSS and online communities within Finland. This has increased the interest in studying the underlying outcomes from the business perspective. Thus, there is a strong interest to find new ways of utilizing OSS and online communities in relation to the available literature and industrial perspective. These new ways of utilizing OSS and online communities can be exploited to achieve (1) open innovation, (2) viral marketing and (3) competing advantages.

First, Chesbrough (2003) states "open innovation is a paradigm that assumes that firms can and should use external ideas as well as internal ideas, and internal and external paths to market, as the firms look to advance their technology". Likewise, OSS and online communities foster co-creation where users together with developers, companies, public organizations and other community members seek new solutions, products, services or business models.
Thus, OSS development methodology proposes a novel way, supported by the available community knowledge and technology, to leverage open innovation within a company and, ultimately, the industry. However, an ICT technology centre Vice President experience is…

…open innovation relations with OSS are hardly seen in the industry.

Moreover, OSS in contrast to proprietary software licenses, ensure that *anyone* can access and freely distribute the software's source to evaluate, modify, and tailor it according to his/her needs. For this reason, OSS licenses are regarded as "viral" licensing models that foster the software propagation. This enables the creation of spontaneous online communities. Thus, the licensing model of OSS can be a mechanism for companies to diffuse and market products rapidly and with low costs.

In relation with the "viral" licensing model of OSS and online communities, SMEs can find a tool to compete with large enterprises. Despite the differences among large enteprises and SMEs resources, OSS and online communities can be utilized to achieve a cost leadership in the provision of software and services, in comparison with large enterprise's proprietary software, often developed in-house. This is the case of the eLearning Finnish market, where an SME that relies on an OSS eLearning platform is the market leader and is achieving continuous market growth.

5.2 Decoupling of Software Components and Software Vendors

A key aspect of OSS is the openness and availability of the source code. As mentioned, this yields different ways of utilizing OSS and harnessing online communities. However, it also results in the decoupling of software components from software vendors. Thus, the openness and availability of the source code avoids vendor lock-in.

In a similar way as the hardware-software decoupling allows users to install software of their choice in any computer. An industry specialist points out that…

…*companies are currently going through the process of decoupling software components from software vendors to become more competitive.*

The elimination of vendor lock-in together with the support of communities results in major implications for the internal processes of companies and competition. The main benefits for a company are:

- Technology flexibility (Waring & Maddocks, 2001)
- Reduce development time (Van der Linden et al. 2009; Kuan 2002)
- New mechanisms to exert market pressure (Välimäki et al. 2004)
- Develop cooperation (Ramirez, 1999; Prahalad & Ramaswamy 2002)
- Potential competitive advantage (Hecker 1999)
- Cost reduction (Van der Linden et al. 2009; Kuan 2002)
- Foster testing, inspection of the code and documentation (Chan & Lee 2004; Cusumano 2004)

These benefits yield from the openness and availability of the source code and the resources concentrated within a community. However, companies are generally unaware of how to integrate OSS and online communities to their business processes and models. For this reason, it is of relevance to formulate a new business model that unleashes the full potential of this new development methodology.

5.3 Online Communities, OSS and Proprietary Software

The interaction among OSS and online community in line with the business model of a company and the delivered proprietary software still remains fuzzy in the Finnish software industry. For this reason, company professionals that have achieved this point out the importance for companies to...

…*understand the general structure and roles of OSS communities while actively participating in all the development phases.*

According to Ye & Kishida (2003), an OSS community structure resembles Figure 3. In the center of the OSS community structure is the project leader, who sets the pace and targets of the community. The project leader is often the founder of the OSS community. On a second level, there is a group of core members that manage the project together with the project leader and actively work on the development. Other community roles are developers (active and peripheral), bug related participants (bug fixers and bug reporters) and OSS users (readers and passive users).

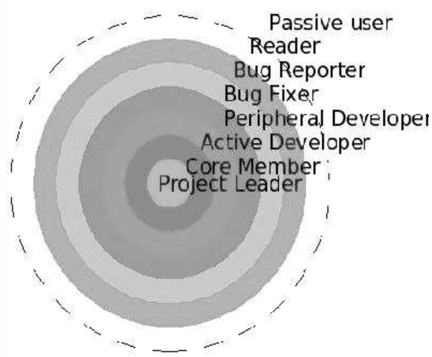


Figure 3. OSS community structure (Ye & Kishida, 2003).

Thus, companies can unleash the full potential of OSS if they can control the community, mainly, through the roles of project leader and the core member group. This enables a company to align its strategic objectives with the community and OSS. An interviewed manager refers to the business case of Nokia and Maemo community where the community, lead by Nokia, develops Maemo Platform embedded in Nokia phones. Figure 4 depicts the integration of online communities, OSS and proprietary software of Maemo case.

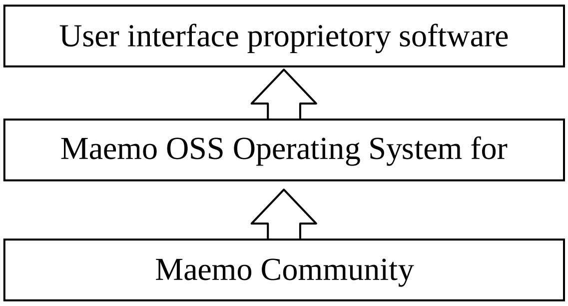


Figure 4. Online communities, OSS and proprietory software – The case of Maemo.

In this way, Nokia is capable of developing user interface proprietary software leveraged by Maemo community and aligned with Maemo platform, built in large parts of OSS components. What is more, Nokia can roll down to the community non-differentiating software and focus on the proprietary user interface as differentiating software. This is in connection with Figure 2 and the OSS strategy pointed out by other authors (e.g. Van der Linden, 2009, and Perens, 2005).

However, this study also found other successful examples of niche companies that, relying on OSS and online communities, do not lead the community or even participate in it. A company's CEO implementing this strategy explains the main disadvantage…

…each new OSS upgrade is seen as a challenge for our IT team as the interoperability with our proprietary software is endangered.

Nevertheless, professionals agree that this strategy brings additional cost savings in terms of memberships and IT personnel.

6 Conclusions

OSS is a natural evolution of the software industry. This new concept refers to a development methodology that, in contrast to proprietary software, provides the source code for the customer, resulting in better software, measured in terms of reliability, usability and interoperability, which often yields shorter development times while, at the same time, enhancing the communication between developers and the community.

The paper showed the growing interest of the Finnish industry towards the dyadic relationship among OSS and proprietary software in connection with online communities. As a result, companies not only take advantage of the community developments on non-differentiating and in-house differentiating software components, but also take advantage of the community knowledge creation and developments in differentiating components.

The dyadic relationships among OSS and proprietary software must be foster and harness by companies through technology and organizational means. Thus, a company shall focus in providing an added value and, in most cases, leading the community to pursue participation and knowledge creation, while the proprietary software aligns with the OSS community developments and the planned roadmap. Furthermore, the next research step is to identify the requirements a business must have to force the company to lead an online community and achieve revenue.

REFERENCES

Brusoni, S., Prencipe, A., Pavitt, K., 2001. Knowledge Specialization, Organizational Coupling, and Boundaries of the Firm: Why Do Firms Know More Than They Make? *Administrative Science Quarterly, 46(4): 597-621.*

Carson, D., Gilmore, A., Perry, C., Gronhaug, K., 2001. *Qualitative Marketing Research*. Sage Publications., London.

Chan, T. Y., Lee, J. F., 2004. A Comparative Study of Online User Communities Involvement in Product Innovation and Development. In *13th International Conference on Management of Technology (IAMOT).*

Chesbrough, H. W. (2003). *Open Innovation: The New Imperative for Creating and Profiting from Technolog.* Harvard Business School Press, p. xxiv

Coar, K., 2006. The Open Source Definition. http://www.opensource.org/docs/osd. 30-11-2009.

Cusumano, Michael A. (2004) *The Business of Software: What Every Manager, Programmer, and Entrepreneur Must Know to Thrive and Survive in Good Times and Bad*. Free Press

Hecker, F. (1999) Setting Up Shop: The Business of Open-Source Software. *IEEE Software (16:1) 1999, pp 45-51.*

Hoch, J., Roeding, R., Purkert G., Lindner, K., 2000. *Secrets of Software Success*. Harvard Business School Press.

Kuan, J., 2002. Open Source Software as Lead User's Make or Buy Decision: A Study of Open and Closed Source Quality. *Open Source Software: Economics, Law and Policy, Toulouse, France.*

McDermott R. (2001). Designing living community events. *Knowledge Management Review*.

Merton, R. K., Fiske, M., Kendall, P., 1956. *The Focused Interview: A Manual of Problems & Procedures*. Glencoe, IL: Free Press.

Pavitt, K., 2002. Innovating Routines in the Business Firm: What Corporate Tasks Should They be Accomplishing. *Industrial and Corporate Change, 11(1): 117-133*

Perens, B., 2005. The Emerging Economic Paradigm of Open Source, *First Monday, vol. 10, no. 10.*

Prahalad, C.K. & Ramaswamy, V., 2002. The Co-Creation Connection. *Strategy & Business, Second Quarter, 27: 1-12.*

Preece, J., 2000. Online Communities: Designing Usability, Supporting Sociability. *John Wiley & Sons*.

Ramirez, R., 1999. Value Co-production: Intellectual Origins and Implications for Practice and Research. *Strategic Management Journal, 20(1): 49-65.*

Stutz, D. (2004). Some Implications of Software Commodification. http://www.synthesist.net/writing/commodity_software.html, 13-02-2010.

Välimäki, M., Oksanen V., 2004. The impact of free and open source licensing on operating system software markets. *Telematics and Informatics, Volume 22 , Issue 1-2.*

Van der Linden, F., Lundell, B,. Marttin, P., 2009. Commodification of industrial software, a case for open source. *IEEE Software.* 26, 4 (Jul. 2009), 77-83.

Waring, T., Maddocks, P., 2005. Open Source Software implementation in the UK public sector: Evidence from the field and implications for the future, *International Journal of Information Management, Volume 25, Issue 5.*

Ye, Y., Kishida, K., 2003. Toward an Understanding of the Motivation of Open Source Software Developers. *Proceedings of 2003 International Conference on Software Engineering (ICSE'03), Portland, OR, 419-429.*

ANTENNA ARRAY WITH ELECTRONIC BEAM STEERING FOR TELECOMMUNICATION APPLICATIONS

Hadi Aliakbarian, Vladimir Volskiy, Guy A. E. Vandenbosch
ESAT-TELEMIC, Katholieke Universiteit Leuven, Kasteelpark Arenber 10, Leuven, Belgium
hadi.aliakbarian @esat.kuleuven.be

Keywords: Electronic beam forming, digital beam forming, beam steering, antenna arrays,

Abstract: Electronically controlled beam steering array antennas highly increase the performance of telecommunication links by improving the link budget. Moreover, digital (baseband) beam steering includes the possibility of having multiple beams simultaneously and of communicating with more than one user at the same time. The basics of digital beam forming are explained in this paper. A 1x2 prototype array with digital beam steering is introduced and measurement results are shown.

1 INTRODUCTION

Beam steering is a very attractive alternative to improve the performance of mobile communication systems and the tracking capabilities of search radars. When the steering is performed digitally new ways are offered to build highly effective wireless communication systems. A conventional antenna array is able to serve users covered by the antenna main beam. To serve selected users in different directions, the main beam can be steered to other sectors using phase shifters coupled to the antenna radiation elements. This technique is called electronic beam steering, or phased arraying. Another solution is to employ several antennas covering the necessary angular sectors. However, adding the beam steering capability improves the performance inside each sector.

The current state of digital technology allows treating the signal received by each array element separately. In practice analog to digital converters (DAC) or digital to analog converters (ADC) are included in the receiving or transmitting channel of each antenna radiating element. Then these signals are digitally processed in a special way using digital devices such as field-programmable gate arrays (FPGA). FPGA's are able to multiply channels and perform processing in parallel. This results in the ability to support several main beams at the same time and/or to form special radiation patterns. As a result each user can be served separately with maximum available efficiency. This is one of the appealing possibilities that digital beam forming introduces for both telecommunication and radar with very low cost.

Within the frame work of the IS-HS-II project funded by the Flemish government, the communication payload for a LEO satellite is being developed jointly by Katholieke Universiteit Leuven and University of Stellenbosch, South Africa. The payload should be able to collect data simultaneously from base stations distributed over several agricultural fields and then transmit them back to the central ground station. It is obvious that to implement the beam steering capability in a large number of ground stations in the field is expensive. Thus it is planned only for the satellite.

In section II, the improvement of the link budget, which is the major motivation for using beam steerable antennas is discussed. In section III, the basics of electronic beam steering are discussed. In section IV, the designed and tested prototype is shown. Fnally the paper is concluded in section V.

2 PROBLEM OF LINK BUDGET

Suppose that we have a satellite flying over a ground station as shown in Figure 1. The long satellite distance from the ground station causes a path loss of more than 153 dB in broadside direction for a satellite placed in a 500 km height orbit, regardless of atmospheric losses. The problem is even more pronounced when the satellite is at grazing angles. Figure 2 shows the path loss for the different relative satellite angles. Thus, the primary calculation of the signal to noise ratio (SNR) based on a simple omni-directional antenna gives a very poor performance in broadside direction and even a negative SNR as the satellite recedes more than 25 degrees, as shown in Figure 3.

In order to have the best performance on the satellite, we have to steer the radiation pattern to the broadside direction as the satellite approaches the station, as illustrated in Figure 1. This can be done either using a mechanically steerable antenna such as a high gain parabolic antenna rotated by an accurate controlled stepper motor or a beam steerable phased array antenna. Obviously, mechanical steering has considerable mechanical drawbacks and it is highly expensive. Thus, today electronic beam steering gradually takes over the steering requirement.

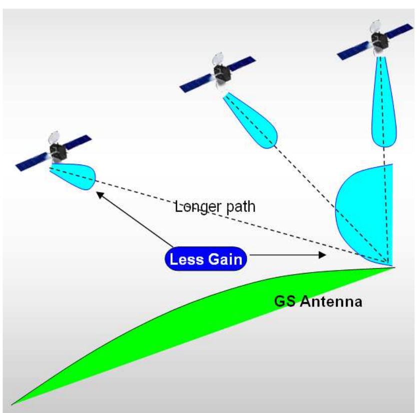


Figure 1: Diagram showing a satellite flying over the base station.

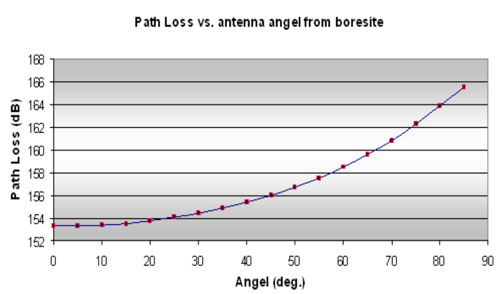


Figure 2: Signal path loss for different angles which satellite is observed.

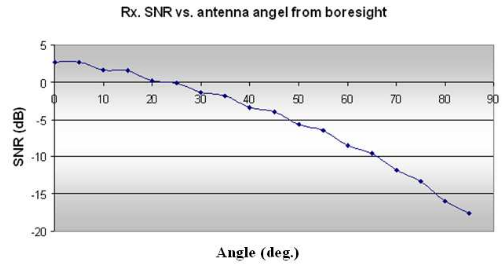


Figure 3: SNR calculated for the link budget.

3 Electronic Beam Steering Mechanism

Electronic beam steering may be implemented when an array has at least two antenna elements. The principle of beam steering is based on applying certain phase shifts on the received signals of these elements, and then summing them together. For example, for the three antennas shown in Figure 4, the incident plane wave (with the angle φ from the horizon) is received at each antenna element with consecutive Δt time differences.,This corresponds to a Δδ degree phase difference in frequency domain, which can be calculated using equations (1) and (2)

$$\Delta t = c \cdot d \cos(\varphi) \tag{1}$$

$$\Delta\delta = \frac{2\pi d \cos(\varphi)}{\lambda} \tag{2}$$

where *d* is the distance between elements, an λ is the wavelength. If the phase difference among signals are compensated using the phase shifter shown in Figure 4, the output signal will be maximized.

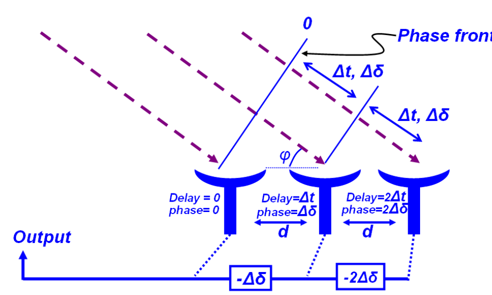


Figure 4: Schematic of a phased array.

This important phase shift may be realized in the RF part, the IF part, or the baseband part of the system. Each solution has some advantages and disadvantages. As a result of the use of RF phase shifters, RF beam forming has several limitations:

primarily high insertion loss, limited phase range, and power consumption. Alternatively, IF beam forming can also be used. It has the same basics as RF beam forming, but the lower frequency has both positive and negative consequences. Circuit design in general becomes easier, components are more readily available, and circuit losses are lower. On the other hand, delay lines become longer and the required mirror frequency rejection usually is a problem (Kashif, 2001), as well as increased inter-symbol interference (ISI).

The capabilities of digital technology have increased considerably in recent years. This has boosted the use of Digital Beam Forming (DBF) compared to Analogue Beam Forming (ABF). However, what we know today as DBF is actually IF band or baseband DBF (DBBF), due to the fact that the phase shift between the elements is applied in these bands. On the other hand, ABF, which today is more frequently realized at RF frequencies, can also be performed at IF or baseband frequencies, namely Analogue Baseband Beam Forming (ABBF). A complete implementation of an 8x8 ABBF array antenna has been done at KULeuven (Aerts, 2009). Therefore, to choose between them, a comparison of Analogue and Digital Baseband Beam forming (ABBF vs. DBBF) has been done recently (Aliakbarian, 2010). The comparison shows that except in special cases such as high bandwidth requirements, DBBF is better. Moreover, an advantage not mentioned in this comparison is the possibility of simultaneous multiple users trac(k)ing, which is highly attractive for our application.

3 Digital Baseband Beam Forming Prototype

The link budget suggests that a 4x4 array antenna, with maximum 12 dBi effective array gain, would be enough for our application on the satellite and for preliminary aircraft tests. But before making the final prototype a 1x2 prototype was implemented with as much space grade components as possible at 2.4 GHz. In the current 1x2 digital baseband beam forming antenna, which is composed of two elements, the signal is first of all received by each element, then amplified by an LNA, and finally directly converted to baseband producing I and Q signals. Both signals again are amplified using a pair of programmable amplifiers, and filtered using band pass filters between 15 KHz and 150 KHz to remove unwanted noise. Then they are digitized and sent to the FPGA board.

In order to realize a baseband phase shifter, either digital or analogue, one can use the schematic illustrated in Figure 5. Both the *I* and the *Q* signal are multiplied with two sine and cosine voltages, which are calculated in the controller PC. Finally the phase shifter signals are summed as the final output signal.

In digital beam forming, the schematic can be implemented in an FPGA's VHDL code, as shown in Figure 6. The necessary phase shifts to both signals are applied and then they are summed together. To calibrate the array and balance the output, the FIFO section in the VHDL code captures 64 shots of the signal and sends them to the controller PC to calculate the calibration coefficient. To apply these calibration coefficients, it is enough to multiply them with the phase shifter voltages, namely the sine and cosine coefficients. If another pair of phase shifters is added to the VHDL code, we can easily have two different beams at the same time without degradation of the performance.

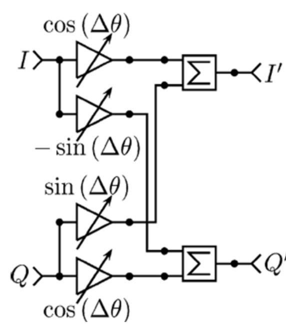


Figure 5: Schematic of a baseband phase shifter.

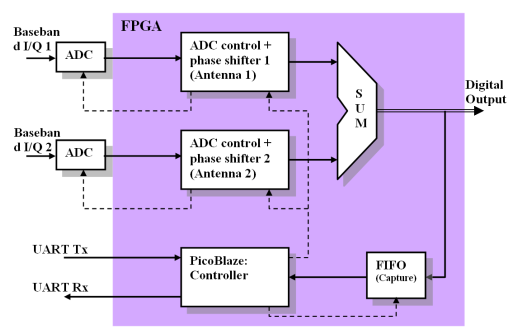
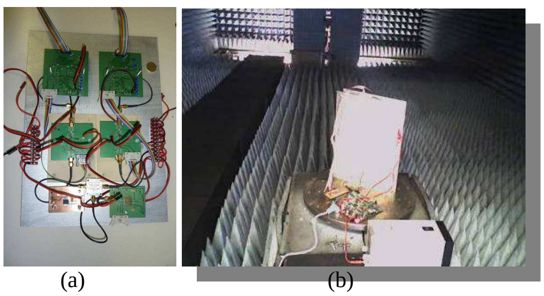


Figure 6: VHDL code blog diagram for the 1x2 prototype.

In figure 7(a) RF, baseband, PLL and regulator boards can be seen. The antenna is then tested in an anechoic chamber. The radiation pattern has a peak at 0^o when there is no phase difference between the elements. Then the radiation pattern is rotated to -

30° by setting the phase shifters and the radiation pattern is measured again. The result of the measurement is shown in Figure 8. As it can be seen the maximum value of the received voltage is now at -30°. There is a considerable sidelobe at +35 degrees which is mainly due to the fact that the element distance is 0.8 λ, which brings the first grating lobe close to the visible range. At the moment, a 4x4 array is being built. This will further increase the gain. The implementation of multiple target tracing is also going to be added.



(a) (b)

Figure 7: (a) RF and baseband Circuit inside the antenna (b) Antenna rotated to -30°.

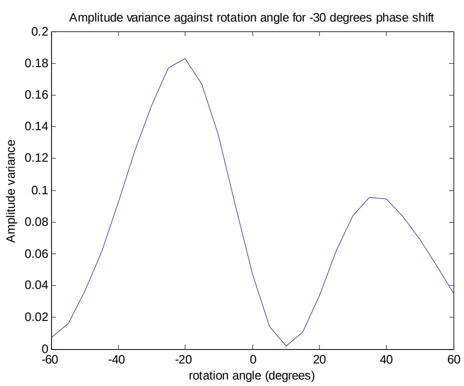


Figure 8: Radiation pattern rotated to 30°.

3 CONCLUSIONS

In this paper, the concept of beam steering is explained. After introducing different methods of beam steering, digital beam forming has been used for the final array to be used in a LEO satellite. The flexibility of digital technology allows us to trace multiple stations simultaneously without performance degradation. A 1x2 prototype of the final digital beam forming array has been implemented and tested and the result has been demonstrated.

ACKNOWLEDGEMENTS

We would like to thank the Flemisch Govenrment, department EWI, for the funding within the framework of the IS-HS project. We also would like to thank Ewald ven der Westhuisen and Dr. Riaan Wolhuter from the University of Stellenbosch, South Africa for their help.

REFERENCES

Kashif, F. M., Qadeer W., and Shah, S. I., 2001. Efficient implementation of quadrature amplitude modulation transmitters. *in IEEE INMIC Technology for the 21st Century*, 2001.

Aerts, W., Delmotte, P., Vandenbosch, G. A. E., 2009. Conceptual Study of Analog Baseband Beam Forming: Design and Measurement of an Eight-by-Eight Phased Array. *in Transaction on Antenna Propagation*, vol. 57, Issues 6, pp.1667 - 1672, 2009.

Aliakbarian, *et al*., 2010. Analogue versus Digital for Baseband Beam Steerable Array used for LEO Satellite Applications, *European Conference on Antennas and Propagation*, 12-16 April, 2010, Barcelona.

SHIPS WITH AN ON BOARD COMPUTER:
An example of how intensive use of TIC's reveals the necessity of a new social organization model.

Xulio Fernández Hermida
Teacher in ETSE de Telecomunicación de Vigo Campus Universitario, 36310 Vigo
xuliofh@uvigo.es

Keywords: ships, social organization,. social organization model, TIC, new era, consciousness, values.

Abstract: The information technologies intensively used on board of the ship can suppose an enormous advance in almost every aspect of the work of extraction of resources of the sea affecting to the way of doing of: Sailors, Captains, Bussiness men and even the Administration. This new way of making things will force us to define a new model of social organization.

The possibilities are enormous. If the performance of the extractive activities increases in a small percentage, we are in front of an individual profit. Typically this improvement in the performance is associated with some costs and this is the game where the market economy works to the perfection. But when the performance has been multiplied by a factor of 10 or 100, and it has been done by all the players in the game; the society, evidently, can not remain playing the same game.

The money and the look for the own profit have been the engines in the industrial era. And this was the game we played along all our lives. It is then very difficult for us thinking in another way to perform. But it is absolutely necessary.

Looking to be rich, we became extremely poor.

1. INTRODUCTION

The information revolution is coincident with, more or less, the end of the XX century. With the information revolution, based on intensive use of ICT, we have come to have tools that multiply by ten or one hundred our ability to interact with the environment. We changed the tools but not the criteria with wich we use them.

In the interaction between technology and society we have given the first step that has been to create the technology. This technology has been created by a few people and distributed to the whole society through the miracle of the market. But nothing has been done in relation to the necessary societal change associated with this change in technology.

For the creation of the technology and his dissemination by all the society we have the tools (universities, companies, society of market...) But we don't have similar tools to overcome the social change.

2. EXAMPLES OF USE OF THE TIC'S IN A FISHING SHIP:

2.1 Autopilot and GPS positioning:

You can plan fishing site from your computer before heading out to sea. The boat moves unattended to the fishing spot and reduces its speed and warns us when approaching the fishing position. We mark the route to launch the fishing arts and the ship will follow it slowly. During the process we can guide the ship by voice through a microphone associated with headphones used by the captain. In the collection of the fishing arts, the boat moves again repeating the route done when launching.

2.2 Sonar technology to locate the banks of fishes:

It is used the sonar of the ship and small sonars on board of buoys. The buoys have GPS positioning and communication with the ship in order to indicate its position and an estimate of the quantity of fishes that there are in its position. These technologies are so cheap that a ship can use hundred of buoys of this kind.

2.3 Localization of thermal fronts in the surface of the sea through satellite images:

From the thermal images from Earth Observation Satellites it is possible to appreciate variations in the temperature of the sea so small as one tenth of degree. Associated with these small changes of temperature, there are banks of fishes that displace with them. The images are collected and processed in research centres and, once located these thermal fronts, transmitted again throgh satellite to the ships that are fishing in the zone.

2.4 Systems for automatic launching of the fishing arts:

Depending on the type of art of fishing, we already have systems for automatic launching, in which case it is not necessary any sailor for that launch. These launching systems together with the automatic pilot and the GPS positioning do that the ship almost can fish by its own. The collecting of the fishes from the fishing art is more complex, and in this process it is still precise human work.

2.5 Artificial vision:

Extracting the fishes from the sea supposes to take them off from the fishing art and afterwards classify them by species and sizes, and putting them in boxes to carry them to the fridge and/or be sold in the fish market (lonja). The systems of artificial vision can be used for this automatic classification and the placing of the fishes in the boxes.

2.6 High accuracy cartography:

With the use of high accuracy cartography and GPS positioning it is possible to avoid launching the fishing arts in zones where the risk that the fishing arts to stick to the bottom is very high. Both these points and the places and optimum routes of launching can be stored in data bases in the ship. This lets to maximize the performance of fishes obtained by launch.

2.7 Data bases in the ship:

Zones of launch, conditions of the sea, season of the year, position of the tide, hour of the day, position of the moon, fished species(in number and size), etc. All of them are data that can remain in the databases of each ship, and that could be joint with the databases of other ships. The knowledge of where to fish can be so precise like the positioning of the roads in a conventional map.

os precios sean más ventajosos.

2.8 Communications, videoconference...

With images or videos of the fishes on board of the ship it can be negotiated the prices of sale and the places of discharge from the same moment in that the fishes are on board of the ship. Discharge of the fish will be done where the prices are more convenient.

2.9 Black box, tachometer (tacómetro):

The administration can force the ships to carry, as the airplanes do, a black box in which it is being registered the activity of the ship. And the content of this black box can be 'Read' by this same administration sometime. The Administration could then, for example, control that the fishing is not being done in fishing reserved zones, or that the ship has not gone out of the fishing allowed waters. It also would serve like defense in front of cases of piracy when fishing ships are kidnapped by administrations of other countries. Another especially interesting use would be like system of alarm in case that the ship need to be helped because it have lost a sailor or in case of shipwreck of the own ship.

3 CONSEQUENCES OF THE INTENSIVE USE OF THE TIC IN ALL THE SHIPS:

3.1 Avalanche effect in the TIC's use

When the use of the TIC is being extended to all the ships it produces an effect of avalanche, in which new profits come to add (or multiply) by the simple massive use. A case of example can be the one lived in the mobile telephony.

Before the mobile telephony mobile terminals, that allowed to some professional enjoy certain mobility, already existed. In these cases it can be said that these terminals are "terminals that depended of only one base station". These terminals are big, high weighed, expensive and its user can not be too much far away from the base station. And only allow him to communicate with his base station.

The first terminal of mobile telephony had the same appearance and the main advantage is that they can connect with any another telephone. And, at the beginning, there was only some very few base stations in some big cities.

The service is very useful, so smaller terminals have been developed (advantage), more base stations have been installed making the zone of service to be greater (advantage), more users buy his mobile terminal and with it there are more people to which one can call although these 'new people with a mobile phone' are not in his office (advantage). The use of sms messages extends although they was initially thought as a residual service (advantage). And an increasingly number of users learn to use this service (advantage). The prices of the terminals remain going down (advantage) and the same do the cost of the communications (advantage). The terminals incorporate video camera (advantage) and allow videoconference (advantage). With the sms messages and the videoconference the deaf incorporate to the society being able to communicate between them in an autonomous way as never before (advantage).

3.2 Coming back to the ships:

When these mentioned systems go extending in use, it will appear other associated elements that will multiply the advantages that we have already quoted. Some examples (that we can try imagine) are:

·That the fish markets (lonjas) get computerized so that they become a central point of concentration of information of captures (that no of real fishes) from where to manage the sales.

·That the ships leave port with a prefixed script for each one so that any one will fish only in the fishing grounds (caladeros) that can be exploded.

·That only a sufficient number of ships leave the port to fish what have decided to be fished.

·That the marine cartographies of the sea floor go improving from data obtained by the equipment installed on board of the ships.

·That embedded sensors buoys be placed in all the zones of fishing so that it can be avoided that a ship to be fishing in zones where the state of the sea is dangerous to be fishing there.

·...

Going out to fish would be almost something more alike to the collecting of grain in zones of cereals or to the vintage in zones of wine. It is fished only what can be fished and what is going to be consumed.

Between all the ships it is fished the necessary and with that, all of us have sufficient fish.

What does not match in this system is the search of the own, individual, profit. An own, individual, profit no longer exist different of the common profit. There exist a win only if this win is a community win.

4 HAVE WE CHANGED FROM SPEAKING OF USE OF THE TIC'S TO SPEAK OF COMMUNISM?

I consider that it is clear that we have to go to another form to organize the work and possibly the whole society. The work and the money are the pillars on which we have built the industrial society of the XX century. The money and his flow is having serious problems and we can see that the work will have to be organized in another way.

Many of the tasks, that until very little constituted a work to be realized by people, are now being realized by machines. And the machines do a work following some computer programs that are being writed in this moment. It is our generation the one who is giving the intelligence to these machines. Giving them the intelligence and the ethical. The work done by a machine can be efficient or inefficient, useful or useless, ethical or no ethical, contribute to the environmental sustainability or shatter resources, help to the welfare of the people or go against us, etc.

As an example of machine doing the work, I remember when spanish public telephony company put those automatic answered machines that simply informed that the mobile number you are calling is not operative in that moment. But if you were calling from a phone box, the call was billed as if you had effectively spoken with a mobile telephone. It results then that you had paid 200 pesetas (more than 1 €) by the information that the called number was not operative. This is an example of work done by a machine and that is not ethical. In the case of spanish public telephon company, there were more cases of this type. I consider this was a dark time!.

There is then an ethical and a set of values that our generation (companies, workers, programmers) are putting into the machines. Machines that are going to be an increasingly important part of our landscape. An increasingly part of the elements with which we will relate and interact along each day.

If these machines have been built with the right values, ethical, favourable to the people, to their dignity and to their realization as human beings in their vital process, we will be building a future with these characteristics. If the values are others, that other values will be our surroundings in the next years.

It is a fact that what we live in each moment (present) is the product of what has been done (by us and by the other) in the past.

In a more transcendental vision, it is said that our life is product of the karma accumulated in our previous lives. And that, depending on the karma purification that we do in this life, that way will be our next lives.

Of the same way, and much more tangible, we can say that the course of a river is consequence of the routes that along the history the waters were choosing to displace in his way to the sea.

Well. We could say that we are now in a period in which these first routes of water drops advancing to the sea are being created. The decisions we are taking today will have an enormous impact in the conception of our future. These present decisions are so important as the decision of this first drop of water in the moment to advance in a direction. The drops that come behind will be very strongly conditioned to follow this same primeval way chosen by this first drop. We have an enormous responsibility in our hands.

It is said that the American Indians, when they had to take important decisions, they took the decision taking in mind "The seventh generation". This is, thinking on how this decision will affect to the children of the children... that will live 140 years after.

In our case the decisions are being taken day by day and by each one of us (although we speak of companies, of organizations, of countries...).

The decisions are took always by the persons. It is crucial that we are conscious of the paper that we are playing and that, as the Indian bosses, we think in the seventh generation.

5 GOING BACK TO THE SHIPS:

The applications that we do can be done so that they maximize profits:

·Find the fish market (lonja) where the prices are higher, get a reduction of personnel on board, reduce fuel consumptions, etc.

·Applications can also be done to improve the life and security of the sailors on board of the ship:

·Systems that avoid the sailor being exposed to fall to the sea water, systems to control of the stability of the ship, systems of alert of bad meteorologic conditions, etc.

Or applications to improve the quality of the fished fish:

·Systems of trazability of the fishes, lines of cold, etc.

Or to improve the control by part of the administration:

·Control of the quantity of fishes in the fishing grounds (caladeros), control of the quality of waters in the surroundings of the ship (bilges (sentinazos), launch to the sea of oils and machinery wastes, fuels, chemical and/or biological wastes, etc.).

When analysed in detail it can be clearly seen that there is not any point not interesting to be considered. And, if we think in the seventh generation, it seems clear that it would be necessary we try to fullfil them all.

However, if we only think in the day of today, the application will serve to the one who pays for it. And the one who pays, if he only thinks in the day of today, he will look for maximize profits. And, like profits in this short vision is directly income least expenses, the application simply will treat to look for fishing the maximum, selling the most expensive possible and reduce to the maximum the costs.

The application only will respect the fishing grounds (caladeros) when the possible fine do not surpass what fished. The application will only invest in security to the sailors when the cost of the compensations to the dead sailors do not compensate what saved in security. The application only will respect minimum sizes of fish when the possible fine do not compensate by the sold. The quality of the fish will be good only in the measure in that this quality can be perceived by the client; it is enough if the fish seems to be good; it does not matter if it is really good or not.

And the TIC's, and all his multiplier effect, can be used to optimize all these aspects.

6 CONCLUSIONS

As in the initial approach, the machines will be doing each time a main part of the work. But is doubtful that, if we build them with these last criteria, the life in the future will be very pleasant. As before, what built today will be what we will live tomorrow.

ACKNOWLEDGEMENTS

I want to thank to my students of the Technology and Society subject I teach in the ETSE de Telecomunicación de Vigo by the interesting debates during the classes and also for the works that they prepare for its presentation in the classes.

Also to my mates of RIGPA, (the Buddhist group in which do meditation) by his interesting observations and fine tunning to the very diverse questions I ask them.

REFERENCES

Dennet D. C. “Me temo que seremos como dioses” 2003, paper in the book “*Haciendo la contra”*, ed. La Vanguardia.

Fritjof Capra “La trama de la vida: Una nueva perspectiva de los sistemas vivos” 1998 ed. Anagrama.

José B. Terceiro, Gustavo Matías “Digitalismo: El nuevo horizonte sociocultural” Ed. Taurusesdigital

DEALING WITH ELECTROMAGNETIC POLLUTION IN POPULATED AREAS

Paula Gómez, Iñigo Cuiñas
Universidade de Vigo, Dept. Teoría do Sinal e Comunicacións, r/Maxwell, 36310 Vigo, Spain
paulagp@uvigo.es, inhigo@uvigo.es

Keywords: electromagnetic pollution, reference levels, electromagnetic exposure

Abstract: Electrical and Telecommunication Engineers live in a society that is concerned about the risk of electromagnetic pollution. A lot of people is against the installation of new radio base stations, and the engineers have to deal with this problem when trying to deploy modern communication networks. A summary of medical reports, as well as an ecological proposal to reduce the electromagnetic pollution are the contents of this contribution.

1 INTRODUCTION

The fear against radio electric stations grows among population in parallel to the deployment of more and more wireless communication systems (Pascandola 2001). Various laws and regulations have been fixed to protect the public health and to control the proliferation of radiation fonts, but this control does not seem to be reflected in the people peaceful.

Several Medicine and Biology research teams have been developed large experiments looking for possible effects of the radio waves on the human health. Up to the knowledge of the scientific community, the health appears to be free of danger when the administrations control the electromagnetic exposure levels and force to maintain them under some restrictions, recommended by the International Commission for Non-Ionizing Radiation Protection (ICNIRP) (ICNIRP 1994 and 1998). However, there exist some lacks of research where no information on hypothetic effects is available: very long term exposures, and others. These are the current objectives of the scientific work, and would be introduced during the presentation of this contribution.

Whereas no new scientific results recommend a constraint in the ICNIRP exposure limits, the reasonable actuation is to maintain these values as a valid recommendation, and to work together with administrations to apply and to track these restrictions. The problem appears when researchers indicate that these limits are valid for adult and health people, whose thermoregulation system works perfectly. This comment opens the doubt on what happens with children, elderly, and ill persons. Trying to surpass such problem some National regulations (Spanish is within) have fixed what they known as “sensitive areas”, where especial care has to be taken to control the radio electric exposure. These areas are mainly kindergartens, geriatrics, hospitals, public parks… in general, places where children and elderly could stand for long periods.

The proposal we present is the use of vegetation barriers to reduce the electromagnetic pollution in such sensitive areas. After large measurement campaigns, attenuations up to 17 dB in terms of received power have been detected. Such values represent an important reduction in the electric field strengths in the shadow zones. The large measurement campaign involving up to seven different vegetation species, and up to ten specimen of any specie, would be described and commented along the presentation.

This attenuation due to vegetation would be an interesting solution to minimize the electromagnetic field strength in places where no users of cellular phone networks or wireless LANs are commonly staying. And these barriers present another advantage: their visual impact is clearly reduced compared to concrete and steel walls that constitute the typical way to obtain radio electric attenuation.

Section 2 summarizes the information concerning medical research. Sections 3 and 4 are devoted of the effect of vegetation in radio electric propagation, both in urban random environments and applied to reduce the electromagnetic pollution.

2 MEDICINE EVIDENCES

The fear to radio electric waves seems to grow in opposite direction to medical research results: the main well determined effect of electromagnetic emissions at mobile phone frequencies is the temperature increase of organic tissues, like the epidermal or the muscular ones. However, the general risks seem to have been exaggerated (Kabat 2008).

Exposure to electromagnetic fields (EMF) is not a new phenomenon, it is ubiquitous and we deal with it since we are born. The natural electromagnetic environment is originated from terrestrial and outer-terrestrial sources such as electrical discharges in the earth's atmosphere and radiation from sun and space. Natural fields commonly have broadband spectrums where random high peak transients or bursts arise over the noise-like continuum background. These natural field levels are orders of magnitude below local field produced by man-made radiofrequency (RF) sources. The everyday use of devices and systems emitting radiofrequency electromagnetic fields is continuously increasing, and so does the number of different artificial man-made RF sources. As technology deployment advances, the fear to possible health consequences coming from a long-time radiofrequency exposure has increased among people, specifically to those effects that may be related with mobile telephony, such as terminals and base stations.

The effects of electromagnetic waves in biological systems are determined in part by the field strength and in part by the amount of energy contained in each photon. Depending on the frequency and energy, electromagnetic waves can be classified in ionising and non-ionising radiations. Ionising fields, like X-Ray or Gamma-Ray, have enough energy to ionize matter, what means that they can break off the chemical links between elements compounding it. On the contrary, non-ionising radiation does not have such energy and so they are not capable of breaking molecular links. This is the case of radiofrequency and microwave systems, extremely low frequency systems, static fields, and several other systems working in bands from several MHz to few GHz. Although non-ionising radiations cannot damage molecular links, they can interact with matter in several other ways. Among their biological effects can be found the heating or temperature rising of organic tissues as epidermal or muscular ones (hyperthermia), the alteration of the chemical reactions or the electric currents induction in tissues and cells.

Most of the damaging effects for health coming from radiofrequency fields above 1 MHz are associated with induced overheating processes. Warming can lead to 1º C upwards of temperature increment, and can cause several physiological responses and a thermoregulatory process, particularly a decrease in the capacity to perform mental or physical duties as the temperature rises. Similar patterns have been confirmed in persons working under extreme heat or being feverish during a long time. Induced overheating can also affect the correct development of the fetus during pregnancy, but only if the increase in temperature is raised over 2 ºC to 3 ºC for hours. Other secondary effects under study are male fertility and sudden appearance of ocular opacities or cataracts.

Until certain limited levels, warming will be cancelled by the normal thermoregulatory processes of the body, and the subject will even not be aware of the increase of temperature. Tissues sensitivity varies depending on the main location of the exposure, and effects are different if overheating is local or affects the whole body, with eyes and testicles as particularly sensitive spots. Furthermore, body heating has diverse impact among particular groups with higher thermal sensitivity, as elderly, children, fetus and people under treatment with drugs affecting their thermal tolerance (SCENIHR 2006).

Below 1 MHz, main effect is the induction of electrical currents in the organism. Exposure to low level radiofrequency fields, insufficient to produce increase in temperature, can alter electric activity of the brain in animals, because they can modify calcium ions. This effect has also been observed in tissues and isolated cells. Other studies suggest that interaction with radiofrequency fields can change cell rate proliferation (effect typically associated with cancer), alter enzyme activity or affect cell DNA. However, neither these effect have not been well demonstrated, nor their health consequences are enough known to be able to restrict human exposure. Field levels of exposure for common population are far lower than the ones needed to achieve a meaningful effect. Up to now, in short-term exposures to common field levels at houses or the environment, no damaging effects have been notified.

The controversy is settled in possible hazardous effects due to radio electric fields below the threshold to produce organism hyperthermia, and when dealing with long-term exposures. Nowadays, scientific community has a broad ignorance over biological effects of the low intensity electromagnetic fields in living organisms, with the special feature that this is a field constantly evolving.

To prevent long-time exposures to radiofrequency fields to have damaging consequences on population, several legislation and directives have been adopted. Recommendations published by different international organizations and commission are based in demonstrated health effects, most of them related with immediately effects coming from short-term exposures, as the stimulation of peripheral nervous and muscles, electric shocks and burns caused by conducting objects, and the generation of high temperatures in tissues because of energy absorption during electromagnetic field exposure.

As various studies seem to show several types of change in cells behaviour, cancer has been the main subject to focus when studying the results on health. Based on the available evidences over health and electromagnetic waves, the International Agency for Research on Cancer (IARC), as a part of the World Health Organization has categorized several agents in five different groups depending on their carcinogen capability (IARC 2006):

Group 1: Agents classified within this group are considered carcinogen for humans. Results are based on strong evidences and analysis during long-time experiments and observations. Examples are asbestos, benzene. Radon, hepatitis C virus, tobacco smoking and tobacco smoke.

Group 2A: It contains agents being probably carcinogen for humans, following conclusions based on animal experimentation. Examples are trichloroethylene, ultraviolet radiation, Diesel engine exhaust, use of sunlamps and sunbeds.

Group 2B: Inside this group, agents are likely carcinogen for humans. Studies are based on evidences in human beings, but there is not possible to assure it because other explanations and reasons may be considered. Examples are magnetic fields (extremely low frequency), work in textile manufacturing industry, zidovudine (AZT), gasoline engine exhaust, coffee (urinary bladder).

Group 3: There are no data enough to classify the agents in this group as carcinogen for humans. Examples are electric fields (static and extremely low frequency), magnetic fields (static), polyethylene, and personal use of hair colour products.

Group 4: The agents here contained are probably not carcinogen for humans. Example is caprolactam.

Some studies, analysing relationships between cancer and radiofrequency transmitters, have been published, most of them referring to leukemia in children and adults (Hocking 1996) (Cooper 2001) (Michelozzi 2002) (Ha 2007). Most of these experiments analyse the distance as an indirect way to measure the exposure. However, results achieved are contradictory and difficult to interpret, mostly because the weak sample of insufficient number of persons, the difficulties to characterize the exposure or the difficult in choosing the geographical areas to compare, among others.

Despite some investigations seem to be able to establish an apparent positive relationship between the leukaemia and the proximity to a base station (Michelozzi 2002), experimental results seem to be inconsistent, and they not show coherence among the different levels of exposure to a radiofrequency field. So, nowadays, it could be said that population exposure to radiofrequency transmitter stations in very weak, and it could be assumed an absence of health damaging effects. Also, all the harmful effects related with this kind of exposure have been identified in studies about mobile terminals more than base stations or in groups highly exposed but in a working scope instead of common civilians.

When a person is exposed to radiofrequency fields, the whole body absorbs energy along time. The specific absorbed rate (SAR) varies along the body, and for mobile terminals, exposure is mostly limited to the part of the head nearer to the antenna of the cell phone. As the field intensity decreases rapidly with distance, user exposure depends on the distance between the cellular and the head. This reasoning has led to several studies relating the use of mobile terminals with the development of intracranial tumours or nervous system tumours (Muscat 2000).

As the fear to electromagnetic emissions has grown among civilians, governments all around the world have adopted precautionary measures to maintain under control health risks in front of scientific uncertainties. Despite World Health Organization has advised the countries do not fix restrictive policies further than the established knowledge, in 1999, WHO was encouraged by countries to have in account the need of applying the Principle of Precaution to evaluate the risks and to be able to adopt a more preventive and proactive focus towards the danger (Kheifets 2001) (WHO 2002).

Among numerous policies promoting precaution most important are:

- Precautionary Principle (PP): This is a risk administration policy applied in circumstances with a high scientific uncertainty (O'Riordan 1994). It reflects the need of acting in front a serious potential risk without waiting for the scientific community to have any result. It is clearly designed to formulate provisional answers to provide responses with higher scientific basis, without expecting until appropriated information is available.
- Prudent Avoidance (PA): It was initially developed as a risk management strategy for electromagnetic fields of energy coming from electric networks (Kheifets 2001) (WHO 2002). In this context, caution was defined as carrying out avoiding actions which imply modest costs. Ever since then, it has evolved and nowadays it implies taking simple, easily reachable, low cost actions to reduce the exposure to electromagnetic fields, even in absence of a demonstrable risk.

Generally, government agencies apply this policy only in new installations, where small design modifications may reduce public exposure levels. However, it is no used to modify already existing installations, because the big cost it usually involves under such circumstances.

3 PROPAGATION IN VEGETATION AREAS

It is well known that vegetation induces additional attenuation to propagation waves, and that this attenuation depends on several factors: frequency, vegetation density, humidity, etc. (ITU 2003). However, this knowledge appears to have slight practical application in wireless network deployment, as coverage seems to be different at arid and vegetation areas when a homogeneous coverage could be expected if the presence of trees and shrubbery would be taken into account when designing. Checking this non homogeneity of the radio coverage is the objective of the measurement campaigns performed in Vigo and Oviedo.

2.1 Environments

The city of Vigo is located in the Atlantic coast of Spain. The population of Vigo is almost 300,000 although the number of people who actually work, study and live in the city may reach half a million.

The measurements in the streets and parks of Vigo were taken during the Spanish national campaign to check the radio electric pollution levels in places where people stay (Min. Pres. 2001) (Min. Cie. 2002), at 2002. The applied procedure was defined by the national legislation, and included a two step method: the first step was an exploration measurement and the second one the exposition measurement.

The equipment used was an electric field strength meter, with a broad band isotropic probe. Concretely, the Narda EMR-300 with the probe type 18 provides the electric field strength in a band from 100 kHz to 3 GHz.

The exploration measurement step consisted on a walk along streets and parks around each mobile phone base station, looking for the places with higher field strength levels. Once five or six of such places were selected, the second step was applied at each location.

The exposition measurement step takes six minutes, averaging field strength data sampled at a rate of 1 second or less. The averaged value was then compared to the limits defined by ICNIRP (ICNIRP 1998), assumed by the European Union (Council 1999). As the National legislation defines “sensitive areas”, as gardens or parks in which children and elderly stay, a lot of measurements at urban vegetation environments were taken during the campaign.

Almost two hundred locations were used for this study, being the 25 % of the measurements gotten at vegetation environments, and the rest in arid or almost arid places.

The Municipality of Oviedo is in the North of Spain. Its current population is about 209,000 people. Oviedo enjoys more than one million square meters of parks and open spaces, the vast majority of which correspond to the large Oviedo parks. This makes the city an interesting environment for the purposes of the experimental work.

The measurement procedure applied during Oviedo campaign was similar to that used in Vigo, with a little difference. The data were taken during the annual campaign to check the electromagnetic pollution in “sensitive areas”: the National legislation, once every base station environment was certified at 2002, obligated to measure yearly in those areas. This means that the surroundings of any park, children garden, hospital or geriatric centre had to be measured yearly. The exploration step of the measurement procedure was then reduced to a short walk constrained to the target environment, instead of the long walk around the base station. The exposition measurement followed the same procedure applied in Vigo: averaging field strength data sampled at a rate of 1 second or less, during six minutes.

The measurements were performed by a portable field meter PMM 8053A with an isotropic electric field probe PMM EP300, covering a frequency range from 100 kHz to 3 GHz.

Almost fifty locations were used for this study, being the 25 % of the measurements gotten at vegetation environments, and the rest in arid or almost arid places.

2.2 Results

The measured results could be classified into two categories: those obtained in wooded areas and those from places without trees. Using this classification, several comparisons between statistics obtained at both kinds of environments are feasible and so that it allows the definition of the different behaviors of both families of theaters. The general trend is that median measured field strengths are larger in arid areas than in wooded areas. Moreover, measured values are more concentrated around the median in groves and parks than in places without trees.

The measurement outcomes at Vigo indicate that the median electric field strength at wooded areas in Vigo is clearly lower than that measured at open areas: from 0.28 V/m to 0.70 V/m. This represents a reduction of 60% due to vegetation, which could represent a reduction in the communication system performance. Another interesting observation is the distribution of the measured values around the median: whereas most of the measurements are closely near their median in presence of vegetation, in zones with no trees the range of values extends to larger numbers. This fact is reflected in the larger extension of the box in the graphic, and in the presence of more outliers when the data have been obtained in open places.

As in the Vigo circumstances, the median electric field strength at wooded areas in Oviedo appears to be clearly lower than that measured at areas with no trees: from 0.30 V/m to 0.67 V/m. This represents a reduction of 55% due to vegetation in terms of received electric field strength, respected to open areas. Moreover, the dispersion of the measured values around their median is larger in arid areas than in wooded ones in Oviedo, which is consistent to the observed in Vigo results.

In general terms, the median field strength in wooded areas appears to be 35% less than in arid areas, in the better balanced case, and 60% lower in the most imbalanced. This clearly different trend between both kinds of areas is confirmed by the mean, which is at least half in groves compared to arid zones.

The standard deviation is visibly larger in arid areas, as a consequence of a wider range of values and their more disperse distribution. Moreover, the range of values is wider in the measurements at open areas than those obtained in wooded places.

The degree of asymmetry of the electric field strength distribution around its mean has been characterized by the skewness coefficients. The distribution is right-skewed in all cases, which could be expected as the range of possible values is limited in the lower values by the sensitivity of the sounder, and the highest measured field strength never reached the saturation level of the equipment.

Finally, a light trend in terms of peakness can be observed by the kurtosis coefficients. A negative kurtosis coefficient, as obtained in the wooded places in Oviedo, indicates distributions where a larger proportion of the values are towards the extremes, that is, relatively "fat" or "heavy" tails compared with a normal distribution. A positive kurtosis coefficient, on the other hand, indicates distributions where the values are bunched up near the mean, i.e., relatively "thin" or "light" tails compared with a normal distribution.

The measured results show that most of the situations the electric field strengths distribution has thinner tails than a normal distribution; and that the wooded environments present relatively heavier tails than the corresponding non vegetated locations at the same city.

The previously presented results confirm that the cellular coverage is less homogeneous as desired in our cities, being reduced in vegetation places due to the presence of trees, which effects appear to be not correctly incorporated in the prediction or planning tools.

4 VEGETATION BARRIERS

An extensive and exhaustive measurement campaign has been developed under actual propagation conditions with five vegetation barrier configurations of different thickness. The selected frequencies for the purpose of this study are those assigned to cellular mobile phone communication systems in Europe (900 MHz, 1800 MHz and 2100

MHz), and in many regions all around the World. The objective of this campaign is to try to demonstrate the effectiveness of using vegetation barriers to reduce electromagnetic pollution, and also collecting samples enough to obtain a mathematic model characterizing the attenuation induced by a vegetation barrier.

The study involved seven vegetation species, both indoor and outdoor shrubs and small trees, installed in five different geometrical configurations. Attenuations up to 9.9 dB have been detected, which supports the proposal of the paper. Therefore, the construction of vegetation barriers seems to be an efficient method to reduce the electromagnetic pollution within sensitive areas.

Narrow band measurements have been used to characterize the effect of the vegetation barriers in the radio channel. The measurement setup is based on commercial equipment used as transmitter and receiver, in co-ordination with an automated linear positioner. Five barrier configurations were considered, involving elements of seven different species. The objective of this campaign was to obtain enough data to establish the relation among accurately attenuation, frequency, and canopy diameter and foliage density of the vegetal specie.

Up to 42 million received power samples were collected during the measurement campaign. At any receiver location, 8001 samples have been measured. These values have been pre-processed to eliminate spurious or outlier measured data and multipath effect due to the height of the antennas. Then, attenuation due to vegetal barriers has been computed by comparing received power in free space condition to that obtained in obstructed line of sight situation.

Results obtained for Irish Juniper and camellia seem to indicate that those species could be suitable to conform barriers providing the attenuation requirements for mobile phone bands at sensitive areas.

A common point in all the configurations and species is that they show larger attenuation levels for vertical polarization than for horizontal polarization. This is probably due to the mainly vertical nature of the trees and shrubs, which should represent a stronger obstacle for the polarization of the same geometry.

Among all the species under study, those with wooden trunk and very dense canopies, as the camellia trees, the Irish junipers and the white cedars, appear to be the most suitable to perform vegetation barriers. Other species offering high attenuation levels are ficus at both polarizations and areca palm at horizontal polarization, probably due to the natural spreading of their leaves in a semi-horizontal orientation, depending on the specimen.

5 CONCLUSIONS

The population fear against electromagnetic installations seems not to be related to what medical research indicates. However, the telecommunication engineers have to deal with this fear when trying to install new radio stations.

This contribution analyses the possible effects of electromagnetic waves on people’s health; but it goes further.

After conclude the attenuation induced by vegetation within the radio channel, mostly in a random way, the authors propose the use of lines of trees to reduce the electromagnetic exposure in some areas that could be considered “sensitive”. Attenuations up to 9.9 dB have been measured in the shadow area of such barriers.

ACKNOWLEDGEMENTS

This work has been supported by Xunta de Galicia (project reference InCiTe 08MRU045322PR), with FEDER confinance.

REFERENCES

D. Cooper, K. Hemmings, P. Saunders. “Cancer incidence near radio and television transmitters in great britain. I. Sutton , Coldfield transmitter; II. All high power transmitters”. American Journal of Epidemiology, 2001. Vol. 153, No. 2, pp. 202-206

Council Recommendation of July 12, 1999 on the limitation of exposure to the general public to electromagnetic fields (0 Hz to 300 GHz), 1999/519/EC, Official Journal of the European Communities, L199/59-70, 30-07-1999.IARC "Complete List of Agents evaluated and their classification". IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. International Agency for Research on Cancer, World Health Organization. January 2006

M. Ha, H. Im, M. Lee, H. J. Kim, B. Kim, Y. Gimm, J. Pack. "Radio-Frequency Radiation Exposure from AM Radio Transmitters and Childhood Leukemia and Brain Cancer". American Journal of Epidemiology, 2007. Vol. 166, No. 3, pp.:270-279

B. Hocking, I. R. Gordon, H. L. Grain and G. E. Hatfield. "Cancer incidence and mortality and proximity to TV towers". The Medical Journal of Australia, 1996

L. I. Kheifets, "The precautionary principle and EMF". World Health Organization, 2001

ICNIRP. “Guidelines on Limits of Exposure to Static Magnetic Fields”, Health Physics, Volume 66, No. 1, pp. 100-106, 1994

ICNIRP. “Guidelines for Limiting Exposure to Time-Varying Electric, Magnetic And Electromagnetic Fields (up to 300 GHz)”, Health Physics, Volume 74, No. 4, pp. 494-522, 1998.

ITU-R, “Attenuation in vegetation”, Recommendation ITU-R P.833-4, 2003.

G.C.Kabat “Hyping health risks”. Columbia University Press. 2008

P. Michelozzi, A. Capon, U. Kirchmayer, F. Forastiere, A. Biggeri, A. Barca and C. A. Perucci. "Adult and Childhood Leukemia near a High-Power Radio Station in Rome, Italy". American Journal of Epidemiology, 2002, Vol. 155, No. 12, pp. 1096- 1103

Ministerio de la Presidencia, Real Decreto 1066/2001, de 28 de septiembre por el que se aprueba el reglamento que establece las condiciones de protección del dominio público radioeléctrico, restricciones a las emisiones radioeléctricas y medidas de protección sanitaria frente a emisiones radioeléctricas, Boletín Oficial del Estado (www.boe.es), No.234, 36217-36227, September 29, 2001 (In Spanish).

Ministerio de Ciencia y Tecnología, Orden Ministerial 23/2002 de 1 de enero, por la que se establecen condiciones para la presentación de determinados estudios y certificaciones por parte de operadores de servicios de radiocomunicaciones, Boletín Oficial del Estado (www.boe.es), No.11, 1528-1537, January 12, 2002 (In Spanish).

J. E. Muscat, M. G. Malkin, S. Thompson, R. E. Shore, S. D. Stellman, D. McRee, A. I. Neugut, E. L. Wynder, "Handheld cellular telephone use and risk of brain cancer". American Medical Association, 2000. Vol. 284, No. 23, pp. 3001-3007

SCENIHR. “Possible effects of Electromagnetic Fields (EMF) on Human Health”. Scientific Committee on Emerging and Newly Identified Health Risks (SCENIHR). European Commission Health & Consumer Protection Directorate-General. Directorate C - Public Health and Risk Assessment. C7 - Risk assessment. July 2006

WHO "Establishing a dialogue on risks from electromagnetic fields". International EMF Project. World Health Organisation.2002

M. Pascandola, “Cell phone lawsuits face a scientific test”, Science, Volume 294, pp. 1440-1442, 16 November 2001

T. O'Riordan. "The Precaution Principle in Environmental Management". Centre for Social and Economic Research on The Global Environment (CSERGE) Working Paper GEC 92-03, 1994.

IDENTITY, LIBERTY AND ACCOUNTABILITY IN THE INFORMATION SOCIETY

A critical assessment of the impact of European ICT policies

Hans Graux

Affiliated researcher at the Interdisciplinary Centre for Law and ICT (www.icri.be), K.U.Leuven, Institute of Problem Solving, XYZ University, Sint-Michielsstraat 6, Leuven, Belgium

hans.graux@law.kuleuven.be

Keywords: identity, anonymity, accountability, information society, data protection, data retention, ICT policy

Abstract: This paper examines the evolution in European policy with respect to identification and individual accountability in electronic communications. Through various examples, we illustrate how the current policy and regulatory framework with respect to online accountability has become much more control oriented over the last decade. This trend reflects the desire for greater security in a post-9/11 world, and supports the professionalization of the digital services market. However, there is also a flipside: a clear tension exists between strict accountability and civil liberties. If not carefully considered, a consistent policy focus on identifying, tracing and monitoring citizens has a stifling effect on society. The ability to identify unlawful modes of expression may also hinder socially useful speech, such as legitimate policy criticisms, the identification of abuses and whistle blowing. In an electronic environment, these modes of expression can be much more effective than was traditionally the case, due to the ease with which communication can spread. The paper examines whether these useful impacts are sufficiently considered in current European policy making, and whether a clear strategy with respect to online anonymity and accountability exists.

1 INTRODUCTION: Identity in the information society

Identity has been a complicated notion throughout the development of human society. It can be approached and defined from numerous angles: as a socio-psychological concept, a cultural or geo-political building block, a static or flexible set of data, a legal notion, an administrative practice, or simply the practical challenge of distinguishing an individual among a set of relatively similar peers (Rowland, 2000; Nabeth and Hildebrandt, 2005).

This paper does not aim to explore this ontology of definitions. Rather, it examines the concept of identity only with respect to its role in the development of the information society, and specifically the importance that has been accorded to identity as an enabler for accountability from a policy perspective.

As we shall see on the basis of the examples below, this is a nontrivial issue. When the technical building blocks of the Internet were being designed, the issue of identity/accountability was given relatively short shrift. The core protocols of the Internet (notably TCP/IP) rely only on technical concepts such as IP addresses and MAC addresses, both of which serve to identify network equipment, rather than the individuals using them (Working Party WP37, 2000).

As a result, establishing or proving identity of individual Internet users can be a laborious, unpredictable and slow process, requiring the cooperation of potentially a large number of stakeholders, including information society services providers, ISPs, courts and law enforcement bodies. In practical terms, this affords Internet users a certain degree of anonymity, not because of any conscious technical, legal or organisational measures, but simply due to the cost and uncertainty of trying to identify an individual on the Internet. In the sections below, we will examine the impact of this issue and some of the main policy responses in the EU over the last decade.

1.1 Identity as a prerequisite for accountability

The practical anonymity of end users on the Internet is however relative, and will only go skin deep for the average user. While it is trivially easy to lie on an on-line registration form to obtain user credentials (username/password) based on false information, Internet use tends to leave traces both locally (on the end user’s system) and remotely (on the servers which the end user has accessed directly or via intermediaries). These traces often enable the identification of Internet users.

The main connecting point to link specific communication to the identity of an internet user lies with his or her IP address: any device which communicates over the Internet requires a valid IP address. Currently, these addresses are typically allocated for a limited period of time by the end user’s ISP. If appropriate log files are kept by the ISP, which indicate which IP address was linked to a specific subscriber over time, then this provides a tentative source of information that can be used to establish some degree of identification and thus accountability, even if the subscriber is not always aware of this. Of course, there are several obvious flaws in this strategy, which will be discussed in greater detail below when examining European data retention obligations.

For most communications via the Internet however, some identifying information is available, either in the form of data provided by the end user himself, or any number of instruments which enable behavioural tracking including ISP logs, payment information for commercial transactions, or so-called ‘cookies’ (small text files which are automatically created and/or updated on the user’s system when visiting certain sites, and which allow site owners to track user preferences or behaviour over time). This establishes a baseline of identifying capability, even in the absence of any infrastructural design choices aimed at facilitating this process (Rowland, 2000).

This situation can be criticised for its lack of transparency towards end users, who are typically not aware of the trail of identifying information that they leave in their wake, irrespective of their preferences (Preneel and Dumortier e.a., 2002). However, a key benefit of this data trail is that it permits identification, and thereby also a degree of accountability for on-line activities. Unlawful exchanges of information (e.g. illegal file sharing, posting defaming or racist comments, downloading illegal pornography) can often be traced back to their source and destination on the basis of IP addresses, server log files, and locally stored user information such as browsing histories.

In this way, the existence of these data sources tentatively linking communication to individuals establishes a baseline of accountability: the average internet user will need to acknowledge that the Internet is not a lawless space in which societal and legal norms are suspended in their entirety. In that respect, accountability can be considered as a primary precondition for a mature, reliable and reasonably secure information society. The possibility of identification and thus accountability can thus logically be expected to be a pillar of European ICT policies.

1.2 Anonymity as a prerequisite for societal freedom

On the opposite side of the spectrum, it should also be acknowledged that the certainty of being anonymous in an electronic setting and therefore unaccountable in certain contexts also has clear benefits, which in some cases have even been recognised through explicit regulations.

There is however no ‘right to anonymity’ under European law, at least not from a positive perspective (i.e. there are no regulations explicitly affirming that such a general right exists). Instead, European law takes a much more balanced approach, by establishing a framework that outlines fundamental freedoms and specifies under which conditions the personal sphere of an individual may be violated.

The most relevant examples of this approach are the right to respect for private and family life (including respect for communications) and the right to the protection of personal data. Both of these are currently enshrined in the Charter of Fundamental Rights of the European Union (articles 7 and 8, respectively), which entered into force on 1 December 2009; other European sources include the European Convention for the Protection of Human Rights and Fundamental Freedoms, which similarly contains an article safeguarding the right to privacy (article 8). These rights are however not absolute, and limitations are possible if they are “provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others” (article 52.1 of the Charter).

Given that they are only formulated at a high and relatively abstract level in these texts, the practical application of these rights in different contexts is usually established through more specific documents. With respect to identity, this principle is nicely illustrated by the Data Protection Directive 95/46/EC, which aims to define internal market rules for the processing of personal data, i.e. information relating to an identified or identifiable natural person ('data subject'). Among other points, the Directive exhaustively stipulates several cases under which personal data can be lawfully processed (article 7). If a specific act of data processing (including e.g. the collection or storage of information which can be used to identify an internet user) does not fall under one of these conditions, then that act is not permitted. Obviously, being an instrument that aims to harmonise the internal market, the Data Protection Directive has little impact on other contexts, notably public security, defence, State security and law enforcement. None the less, this approach creates what might be considered a regulatory bias towards anonymity (rather than an actual right to the same), at least with respect to internal market activities: any such activity requiring or resulting in the identification of a natural person requires an explicit justification to ensure its legitimacy.

This requirement in principle for a justification to identify Internet users has a number of societally beneficial side-effects. By establishing a baseline bias towards anonymity (without going so far as to create a right to anonymity), other fundamental rights are stimulated, including notably the freedom of expression and information (article 11 of the Charter). If individuals have a relative certainty that their opinions and expressions cannot be traced back to them, they will enjoy a greater freedom to voice and debate their concerns, and to seek information on topics which interest them personally for whatever reason. In this situation, the existence of relative anonymity serves as an enabler to critical thought and personal development, as can be seen every day on social networks and on-line discussion forums: while much on-line expression/information consumption in these environments is superficial and without much cultural or societal merit, it is clear that individuals feel much more free to express their political and personal opinions, to seek out peers with similar interests and lifestyles, or even to seek help in difficult situations such as marital difficulties or depressions which may be harder to discuss in a less anonymous setting (Working Party WP6, 1997). As formulated in the US in McIntyre v. Ohio Campaign Commission: “*Anonymity is the shield from the tyranny of the majority. … The objective of the First Amendent is to protect unpopular individuals from retaliation - and their ideas from suppression at the hand of intolerant society*”. In that sense, a certain degree of anonymity is essential for the free development of a critical society.

This same consideration has also driven many countries to adopt anonymity rules in specific contexts, including notably whistleblowing rules (protecting those who anonymously reveal specific wrongdoings in an organisation or group of people; see Working Party WP117, 2006) and shield laws (allowing journalists to ensure the anonymity of their sources). In both cases, the possibility of ensuring anonymity is crucial to allow the source of the information to step forward. In the absence of this assurance, the source may instead prefer to remain silent, which can be detrimental when the facts to be revealed relate to a matter of public interest.

The examples above (anonymous expression on blogs, anonymous access to sensitive materials, anonymous whistleblowing and shield laws) illustrate that there is a societal interest in ensuring the possibility of anonymity. If no possibility of anonymity exists, useful modes of expression may be stifled, ultimately harming society as a whole. Clearly, it is important for European ICT policies to seek a delicate balance between permitting anonymous communication and ensuring accountability in case of abuses.

In the sections below, we will examine the policy responses provided to this challenge at the EU level.

2 Identity and anonymity within EU ICT policy: a historical note

The natural tension between the need for anonymity and accountability is of course not new, or even unique to an on-line context. Before looking at current EU ICT policies, it is worth examining how this issue has been assessed historically. As we shall demonstrate with a series of examples, the possibility of anonymous communication was considered of pivotal importance in the earlier days of ICT policy, notably up to around 2000.

In the sections above we have already argued that the EU does not acknowledge any general right to anonymity as such, but that it relies instead on its interpretation of fundamental rights in different contexts (notably the right to protection of privacy and data protection) to provide a certain degree of anonymity. As a result, the opinions issued by the so-called Article 29 Working Party (Working Party on the Protection of Individuals with regard to the Processing of Personal Data) are instructive, since they contain authoritative (but not binding) guidelines on how European data protection laws should be interpreted. The Working Party consists of representatives of the independent data protection supervisory authorities established in each Member State, as well as the European Data Protection Supervisor.

Recommendation 3/97 on Anonymity on the Internet (Working Party WP6, 1997) is possibly the most explicit in this respect, as it directly addressed the balance between anonymity and accountability in an on-line environment. The Recommendation stressed that there was no need for a unique approach tailored towards on-line communication, as the same challenges were encountered in an offline context as well, mentioning as examples the use of letter and parcel post, the telephone, and newspapers. The Working Party noted that "the ability of governments and public authorities to restrict the rights of individuals and monitor potentially unlawful behaviour should be no greater on the Internet than it is in the outside, off-line world. The requirement that restrictions to fundamental rights and freedoms be properly justified, necessary and proportional in view of other public policy objectives, must also apply in cyberspace." Thus, the use of more extensive identification and monitoring obligations than in an offline environment was explicitly rejected. This rule was also echoed in policy statements of the time, including the 1997 " Bonn Ministerial Declaration ("Ministers recognise the principle that where the user can choose to remain anonymous off-line, that choice should also be available on-line"; "Ministers urge industry to implement technical means for ensuring privacy and protecting personal data on the Global Information Networks, such as anonymous browsing, e-mail and payment facilities".) Clearly, in 1997 European ICT policy makers explicitly recognised the legitimacy and benefits of anonymous communication, including anonymous Internet access, anonymous web surfing, anonymous e-mailing, and anonymous discussions.

The Working Party released a much more extensive opinion in 2000, entitled "Privacy on the Internet - An integrated EU Approach to On-line Data Protection" (Working Party WP37, 2000). This opinion applied the aforementioned principles to a number of key information society services (web surfing, e-mail, FTP, newsgroups, etc), in each case indicating how anonymising technologies can improve privacy protection for the end user. Anonymity is again repeatedly stressed as the rule, rather than the exception: "If they are not anonymised, data on searching and surfing on the Internet should not be kept once the Internet session has finished" (Working Party WP37, 2000, p. 49)

Outside of EU bodies, the Council of Europe similarly issued Recommendation no. R(99) 5 for the Protection of Privacy on the Internet in 1999. This recommendation explicitly recognised "the need to develop techniques which permit the anonymity of data subjects and the confidentiality of the information exchanged on information highways while respecting the rights and freedoms of others and the values of a democratic society", and provided a series of guidelines for the protection of individuals with regard to the collection and processing of personal data on information highways. These guidelines included recommendations for end users to use encryption and other anonymising technologies, and to use anonymous internet access ("by using public Internet kiosks or pre-paid access and payment cards") whenever possible. Similarly, ISPs and other service providers were recommended to inform subscribers "about the possibilities of accessing the Internet anonymously, and using its services and paying for them in an anonymous way […]", while urging them to "[d]esign your system in a way that avoids or minimises the use of personal data."

Generally, anonymity of Internet users was considered a positive and desirable possibility in all of these texts. While each of them also contains cautionary notes that legal restrictions may apply which make true anonymity impossible or undesirable for some contexts, they all recognises the legitimacy of remaining anonymous to the same extent as in an offline environment.

It is important to stress that this recognition is not limited to policy documents and opinions, but that it has also been reflected in several regulatory texts. The crucial role of the Data Protection Directive has already been mentioned above. A separate legal framework has been established to govern data protection in the context of electronic telecommunications. As this framework was revised in 2009, its contents shall be discussed below in the assessment of current European ICT policy.

The importance of anonymity has also been explicitly recognised in regulations that do not focus on data protection. A clear example is the eCommerce Directive 2000/31/EC. This Directive provides a basic regulatory framework for information society service providers, and does not focus explicitly on data protection or anonymity. None the less, recital 14 of the Directive explicitly notes that “this Directive cannot prevent the anonymous use of open networks such as the Internet”, thus maintaining the position that anonymity should remain a legitimate option.

A second known example of a regulation supporting anonymous (or rather: pseudonymous) communication is the eSignatures Directive 1999/93/EC. Like hand written signatures, electronic signatures are really only meaningful if they can be attributed to a specific signatory (i.e. the signature has to be somehow linkable to a signing entity). Indeed, Article 2.9 of this Directive notes that the electronic certificates which link the signature to a signature must confirm the identity of that person. However, this is not a strict identification obligation: Article 8.3 of the Directive notes that Member States may not prevent a pseudonym being included in the certificate instead of the signatory's name. Thus, the eSignature Directive actually does not impose a universal identification obligation, but rather a linkability obligation: receiving parties must be able to determine at least by which identity the signatory has chosen to present himself, but not necessarily who the signatory is. This allows signatories to leverage the benefits of eSignature technology (namely to authenticate specific information and to bind it to a chosen persona), without needlessly impairing their privacy.

Other regulations introduce specific identification/transparency obligations in contexts where this is necessary to improve trust or accountability (Preneel and Dumortier e.a., 2002). This is includes the eCommerce Directive, which requires information society service providers to identify themselves (Article 5), and the Data Protection Directive, which contains a similar obligation for the entities responsible for the processing of personal data (Article 10). However, neither one of these obligations apply to Internet users, who have no general identification obligation under any EU regulation.

All of the mentioned regulations, opinions and policy documents thus take a similar position: there is nothing so unique about the Internet that it requires abandoning the possibility of anonymous communication. Rather, the same freedoms available in an offline environment should also be available in an online context.

3 Identity and anonymity in current EU ICT policy: examples and trends

Looking at more recent European policy documents and regulatory initiatives, a decidedly different picture emerges. Whereas civil liberties (including the possibility of anonymous communication or even anonymous payments) were considered an important policy goal ten years ago, various changes and evolutions have virtually eliminated such ambitions from the political agenda.

Key drivers behind this development have been the desire for greater security in a post-9/11 world, and the professionalization of the digital services market, both of which require the reliable identification of relevant stakeholders and the ability to take effective action in case of wrongdoings or abuses.

We argue that the current policy framework has become much more control oriented, and aims to support accountability at every level, in direct contradiction of the positions taken a decade ago. These ambitions can be seen in various ICT related policies, including with respect to electronic communication, e-business, intellectual property protection, e-government, e-payments and cybercrime.

Possibly the best illustration of the change in European policy towards online anonymity is offered by the changes in the telecommunications framework: firstly the various iterations of the ePrivacy Directive, most recently amended by the Citizens Rights Directive, and secondly the Data Retention Directive.

3.1 ePrivacy/ Citizens Rights Directive

The European Union has implemented a separate framework to govern data protection in the telecommunications sector. The first incarnation of this framework was Directive 97/66/EC. In keeping with the policy preferences at the time, Article 6 of this Directive noted that traffic data must be erased or made anonymous once the communication has been completed, with Recital 19 stressing once again that Member States should "encourage the development of telecommunications service options such as alternative payment facilities which allow anonymous or strictly private access to publicly available telecommunications services". The Working Party too supported the importance of anonymous communication through its Recommendation 3/99 (Working Party WP25, 1999), indicating that traffic data should not be kept only for law enforcement purposes and that national laws should not oblige telecommunications operators, telecommunications services and ISPs to keep traffic data for a period of time longer than necessary for billing purposes (Goemans and Dumortier, 2003).

This Directive was replaced in 2002 by Directive 2002/58/EC. This revision retained the same basic approach with respect to traffic data, and added additional protective mechanisms for location data, i.e. data indicating the geographic position of the terminal equipment of a user of a publicly available electronic communications service. However, this revision also introduced an exception to the obligation to erase or anonymise traffic data: Article 15.1 of the Directive explicitly noted the right for Member States to adopt legislative measures to restrict the scope of the general obligation to erase or anonymise "when such restriction constitutes a necessary, appropriate and proportionate measure within a democratic society to safeguard national security (i.e. State security), defence, public security, and the prevention, investigation, detection and prosecution of criminal offences or of unauthorised use of the electronic communication system […]. To this end, Member States may, inter alia, adopt legislative measures providing for the

retention of data for a limited period justified on the grounds laid down in this paragraph". Thus, a regulatory window was opened to permit general data retention obligations, despite the fact that this would contradict earlier policies that online environments should not be confronted with more stringent identification obligations than offline ones, and despite explicit and repeated cautions against the potential undesirable consequences of such an approach from the Working Party.
Finally, this Directive was amended by the Citizens Rights Directive 2009/136/EC, which however did not impact the data retention possibility introduced in 2002. By that time, a general data retention obligation had already become a reality via the Data Retention Directive.

3.2 Data Retention Directive

The Data Retention Directive 2006/24/EC was adopted in order to avoid divergences in national approaches with respect to data retention, as noted in Recitals 5 and following of this Directive (Taylor, 2006). Repeatedly referring in the Recitals to (then) recent acts of terrorism and the need to ensure that sufficient data would be available to law enforcement bodies, the Directive introduced an obligation for European telecommunications service providers to register and retain certain essential information in relation to electronic communications between their customers. This obligation applies to fixed and mobile telephony, Internet access in general, e-mail and Internet telephony. Retention periods should be set by the Member States between 6 months and 2 years.

Thus, all electronic communications via public electronic communication networks between EU citizens were made subject to a blanket logging obligation, irrespective of any indications or suspicions of wrongdoings. In doing so, the Data Retention Directive embodies a reversal of prior European policies with respect to anonymity on the Internet: rather than maintaining the basic principle that the same possibilities of anonymity should be available on-line and offline, the Data Retention Directive seeks to ensure that tools are always available to eliminate this possibility altogether.

Obviously, this change of policy has not been without controversy. The Working Party already strongly criticised the principles and approach of the draft Directive in 2005 (Working Party WP113, 2005), noting that "(t)raffic data retention interferes with the inviolable, fundamental right to confidential communications", and repeated its criticisms after the adoption of the Directive, along with a set of recommendations to minimise the privacy impact of the Directive (Working Party, WP116).

Furthermore, the Data Retention Directive sets a dangerous precedent. Once it has been accepted that all electronic communications can be logged on a continuous basis for up to two years on the grounds that some of it may be unlawful (as has been done through the Data Retention Directive), it is not unthinkable that this line of reasoning would be extended to other contexts as well.

Indeed, proposals in this direction have already been made. On 23 June 2010, Declaration 29/2010 was adopted by a majority of 371 European Members of Parliament, calling for the establishment of a European early warning system (EWS) for paedophiles and sex offenders. The Declaration in fact calls for an extension of the scope of the Data Retention Directive to cover on-line search engines, requiring them to store search histories for individual users for the same period of time as telecommunications service providers (i.e. 6 to 24 months).

The Declaration is baffling in many ways. One of the justifications invoked to defend the proportionality of the Data Retention Directive was that the retention obligations did not relate “data that are the content of the information communicated” (Recital 13, Article 1.2 and 5.2). An obligation to store search histories would accomplish precisely the opposite, as search words are by definition specifying the content that the user is looking for. In addition, it should be noted that the Data Retention Directive applies to providers of publicly available electronic communications services and of public communications networks, neither of which seems to include search engines. Thus, it seems impossible to integrate such an obligation into the Data Retention Directive, as had already been affirmed by the Working Party in 2008 (Working Party WP148, 2008, noting that the Data Retention Directive did not apply to search engines, and stressing that search histories should not be retained once they no longer serve the specified and legitimate purpose they were collected for, with a maximum of 6 months being provisionally proposed.

Clearly, the approach taken by the Data Retention Directive has worrying implications. While strongly supporting the possibility of identification and therefore accountability in case of abuses, it also reduces the possibility of anonymous communication for the average user in a way that seems manifestly disproportionate when compared to the possibilities which exist in offline communication, where casual anonymous conversations remain possible. Indeed, if the Data Retention Directive is considered proportionate, there seems to be no barrier to introducing similar logging obligations in an offline context, including e.g. mandatory registration of any communications via paper mail or even person-to-person communications, and continuously logging personal whereabouts. The increasing prevalence of location-aware mobile communications devices will likely make such an approach technologically viable in a matter of years. It is however very questionable if this development is still compatible with the traditional European notion of privacy, and especially with the principle that exceptions to fundamental rights should be “not excessive in relation to the legitimate needs and interests which have occasioned it” (Goemans, Dumortier, 2003).

4 A role for technology?

4.1 Technology as an enabler for anonymity

The importance of technology as an element conducive to enabling or supporting the privacy of Internet users has long been known and emphasised in EU policy documents. Indeed, much of the Working Party’s 2000 Working Document on Privacy on the Internet (Working Party WP37, 2000) was dedicated to various technologies which could be used to improve on-line anonymity, including direct references to specific products or services that could be used for anonymous surfing, mailing and participation in on-line discussions. Indeed, there is even a limited regulatory basis in the EU for the use of technology to protect privacy (although not necessarily anonymity): Article 17 of the Data Protection Directive explicitly introduces an obligation for those who are responsible for the processing of personal data to implement “appropriate technical and organizational measures to protect personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorized disclosure or access, in particular where the processing involves the transmission of data over a network.” This provision can be considered as a partial regulatory manifestation of the privacy-by-design paradigm *avant la lettre*, i.e. the principle that privacy protection should be embedded in the design specifications of technologies used for the processing of personal data (Cavoukian, 2009).

The development of privacy enhancing technologies is also stimulated at the European level through publicly funded research initiatives, such as the projects PRIME (Privacy and Identity Management for Europe, https://www.prime-project.eu/) and its successor PrimeLife (http://www.primelife.eu). Both of these aim to conduct research and create demonstrators, pilot implementations and toolboxes that support anonymous or pseudonymous communications on-line. As a result, some of the project results negatively impact the effectiveness of the Data Retention Directive.

An example of this is the OnionCoffee software module, which was developed within the PRIME project. This software is a Java implementation of the specifications of the well known anonymous overlay network TOR (http://www.torproject.org). TOR (and the OnionCoffee implementation) allow Internet users to hide their on-line activities by re-routing their communications through various nodes before reaching their destination, i.e. through other systems within the TOR network (Preneel and Dumortier e.a., 2002). In this way, participants in on-line communication would not be able to directly identify each other: each participant would be able to identify which intermediary node in the TOR network they would be communicating with, but not the node that originally requested or sent the information. By increasing the number of intermediaries and rotating them continuously, it becomes increasingly harder for an observer of the system to reconstruct who originally requested or sent specific information.

In that sense, TOR (and a number of similar anonymous communication networks) is an excellent application of the vision of anonymous electronic communication formulated ten years ago. Participants are able to communicate without leaving much in the way of traces, thus ensuring that they have similar freedoms on-line as offline. However, this also means that TOR runs directly contrary to the goals of the Data Retention Directive: by routing data traffic through intermediary nodes, it becomes progressively harder to obtain logs that allow each subscriber to be identified. Furthermore, if even one such intermediary used a system which was not subject to logging, the accountability chain is broken. It is odd that European policies in this area support systematic logging of electronic communication via regulations on the one hand, and also support the development of technologies to thwart the effectiveness of these regulations on the other. This can be seen as an indication that there has not yet been sufficient debate in the European

Union to arrive to a clear consensus on where policy makers stand on the importance of online anonymity and accountability, leading to a multitude of initiatives which are logically hard to reconcile.

4.2 Technology as a policy challenge

The section above highlighted that European policies with respect to anonymity and accountability are incongruous; in the sense that there is no clear vision of where/when anonymity is permissible or desirable. In addition to this lack of a policy vision, an additional problem is presented by the fact that there is a strong gap between the assumptions on which the Data Retention Directive was built and on the realities of electronic communication.

The Data Retention Directive imposes a general logging obligation on the grounds that this is likely to provide useful information to investigate 'serious crime', without specifying what this concept might entail. However, it does not resolve other crucial challenges, including notably the following:

- There is no indication of how service providers such as ISPs are expected to implement these obligations, or of who will pay the cost. It is thus unclear how reliable information logged in this way will be in practice.
- Generalised logging may cause specific difficulties when applied to professions with specific privilege or discretion rules, such as lawyers, doctors, and investigative reporters. Removing the possibility of anonymous communication may prove to be stifling to the availability and quality of their services.
- With the current internet design, specific participants in a communication cannot be reliably identified through logging. The only element that can be logged are certain technical characteristics (such as IP addresses/MAC addresses) of the network equipment being used. Thus, logged information can at best be used to identify specific network equipment; not individual users.
- The reliability of this technical information is suspect, since it can be easily spoofed (falsified). Furthermore, the Directive ignores the reality of the very large number of Internet access points which are intentionally left open to public use (e.g. in hotels, bars, conventions, etc), or which have been left unprotected by technologically ignorant or careless users, or which use encryption protocols that are easily hacked by knowledgeable criminals. Clearly, in this context, logged information can at best be considered an extremely unreliable primary indicator of any wrongdoings, which will need to be supported by substantial additional evidence for the purposes of law enforcement.
- In that respect, it is interesting to note that the use of encryption technologies is free in the European Union, making it easy for individuals to mask specific (lawful and unlawful) communications or content. This is likely to make it increasingly ineffective in the future to target digital information for analysis.

While the EU has not yet implemented specific strategies to plug these holes in the electronic surveillance net, at the national level several Member States have already taken action to remedy some of these weaknesses. With respect to encryption, both the UK and France have adopted regulations making it a criminal offense to hand over encryption keys to law enforcement bodies (respectively via the Regulation of Investigatory Powers Act, part III, and via Article 434-15-2 of the French Penal Code). In both cases, refusal to give up encryption keys can lead to imprisonment of up to five years.

Similarly creative approaches have been implemented to address the issue of open or badly secured access points, with a well know example being the French HADOPI Act (*Loi favorisant la diffusion et la protection de la création sur Internet*) of May 2009. This law aims (among other points) to more effectively combat on-line infringement of copyright protected materials, an area which suffers heavily from the present lack of accountability on the Internet. The Act is based on the so-called 'three strikes' principle, allowing the accounts of Internet subscribers to be suspended after three incidents of copyright infringement. However, it was already noted above that linking an IP address of a suspected infringer to a specific individual is a costly and uncertain process. The HADOPI Act addresses this issue by emphasising the accountability of the subscriber of the relevant Internet account, rather than examining who is factually guilty of misbehaviour. As part of the three strikes process, subscribers are notified of infringements supposedly occurring via their accounts, first by e-mail (strike one) and then via registered mail (strike two). If incidents are still repeated an appropriate time after these notifications (strike three), the account is shut down for up to one year. In this way, accountability is shifted to subscribers whose accounts were used to commit infringements, irrespective of the question whether they are actually responsible. Obviously, provided that this approach is considered acceptable, the same mechanism could spell doom for initiatives such as TOR, since there is no reason why such vicarious liability regimes could not also be applied to any participant in an anonymous network whose node was used for unlawful behaviour.

5 CONCLUSIONS

The European Union has been alternately lauded and chided for placing a significant value on privacy and data protection, as illustrated by the explicit qualification of privacy and data protection as fundamental rights for European citizens, and as witnessed by the Data Protection Directive's status as a uniquely encompassing and strict framework for the legitimate processing of personal data. However, in recent years the EU's policies with respect to identity and accountability are characterised by a certain degree of institutional schizophrenia: while continuously reiterating the importance of fundamental rights such as privacy and data protection, the initiatives listed above are actively working towards a monitoring society in which online activities are continuously logged.

These developments seem to illustrate an increasing lack of balance between the fundamental rights to privacy and data protection on the one hand, and the need for accountability and societal security on the other. Initiatives like the Data Retention Directive suggest that European regulators are much more willing to sacrifice a significant measure of individual privacy in the face of what has been labelled the politics of pre-emption (Morariu, 2009).

Furthermore, there appears to be a large and growing gap between the European legal framework with respect to online accountability and the technological and practical realities in the field. When looking at the Data Retention Directive, the regulatory framework contains the major goals to be achieved in the Member States (as per the general philosophy behind every directive), but very basic questions such as costs and efficiency are given no consideration, leading to huge implementation challenges in practice (Kosta and Valcke, 2006).

Indeed, early implementation efforts have met with strong constitutional objections. In

Romania, the Constitutional Court noted in 2009 that "the regulation of a positive obligation that foresees the continuous limitation of the privacy right and secrecy of correspondence makes the essence of the right disappear" (Renaudière, 2010). Similarly, the German implementation act (the *Gesetz zur Neuregelung der Telekommunikationsüberwachung* of 21 December 2007) was suspended by the German constitutional court on 2 March 2010, on the grounds that it contained insufficient guarantees with respect to the secure storage of logs, lawful use of the data, and because the law could "cause a diffusely threatening feeling of being under observation that can diminish an unprejudiced perception of one's basic rights in many areas." Similar implementation difficulties have been reported in other Member States as well (Renaudière, 2010), and it seems that the Directive’s implementation deadline (15 September 2007, or up to 15 May 2009 for Internet services) has not been effective in ensuring a consistent application of its provisions.

If data retention would be considered necessary to ensure accountability in an online environment, one might wonder whether a more privacy friendly approach would not have allowed for a more balanced outcome based on privacy by design principles, e.g. by requiring the implementation of systems which allow revocable anonymous access to the Internet (Preneel and Dumortier e.a., 2002). As it stands, the European framework offers very little guarantees with respect to the proportionality, reliability and security of logged information. Globally, the Directive focuses very much on increasing the amount of available information, but not on ensuring the usability in practice of this data. As a result, the approach taken by the Directive implies a major economic burden for society and privacy risk to its citizens, with a benefit that can best be described as uncertain.

Of course, these reservations should not be taken to suggest that current regulations would result in a Big Brother society, in which all electronic communication is monitored. For now, the content of communications is not targeted by European rules. In addition, numerous technological solutions are available to improve the anonymity of European citizens on the Internet, including the use of encryption. Finally, practical barriers impede the utility of the current data retention approach: open access points are still widely available, some of which are arguably not covered by European rules.

Yet, for each of these issues there are initiatives, plans or national regulations which aim to eliminate them. The extension of the Data Retention Directive to cover search engines would represent a move towards logging communication content as well. Privacy supported by encryption can be fought by measures such as mandatory encryption key escrow or criminalising the refusal to hand over encryption keys. And the issue of open or badly secured access points can be addressed by simply holding subscribers accountable for any wrongdoings through their connections, as is currently already the approach with three strikes approaches. In the longer term, an increasingly strict regulatory attitude might be able to fully eliminate anonymity in an electronic environment, or at least to ensure that serious sanctions are applied to those who attempt to conceal their activities from any authorised monitoring bodies.

All of these measures would likely have seemed unacceptable to the policy makers of 10 years ago, who lauded the benefits and opportunities offered by anonymous electronic communication. Yet today, each of these measures features prominently in policy discussions. In fact, this evolution may be hard to stop, as there is a clear tendency for such regulations to build upon each other, which each building block naturally suggesting an expansion towards the next. For instance, while the Data Retention Directive may allow logging of communications, optimal use of this information requires that the contents of those communications can ultimately be determined and deciphered. Hence, the logic of the Data Retention Directive naturally suggests expansion into monitoring content, or at least in the need to control the use of encryption technologies. If those next steps are not taken, society might adapt by adopting encryption on a large scale, and the Data Retention Directive would become largely ineffective. Thus, imposing ubiquitous monitoring restrictions creates a vicious cycle which requires continuous extension until all routes towards anonymity are eliminated, ultimately culminating in a ubiquitous surveillance society (McIntyre, 2008). This expansion of supervision powers over different social spheres has been aptly referred to as a ‘superpanopticon’ or ‘electronic panopticon’ (Gordon 1987, cited in Morariu, 2009): an information society with invisible guards, who are capable of monitoring all online interactions, without any way for citizens to determine whether or not any given exchange is actively being observed. Clearly, this makes electronic communication much less attractive whenever discretion or confidentiality are needed.

In the introductory sections of this paper, we have argued that a degree of anonymity is a prerequisite for true freedom of expression, and a crucial enabler to critical thought and personal development. Over the last decade, this notion has increasingly been abandoned in European ICT regulations, which are gradually driving a migration from an open communication infrastructure towards a tightly monitored one. Once initiated, this trend is difficult to break. In this context, it can only be hoped that European policy makers and law makers will rediscover the importance of anonymity in the future.

REFERENCES

Bonn Ministerial Declaration - *Declaration of the European Union Ministers on Global Information Networks: Realising the Potential*, 1997; http://www.gandalf.it/netmark/attach/bonn_en.htm

Bonnici, J.,2007. *Recent European Union Developments on Data Protection…in the Name of Islam or 'Combating Terrorism'*. Information & Communications Technology Law 16, 161-175

Cavoukian, A., 2009, *Privacy by Design*, IPC, Ontario.

Claessens J., Diaz C., Goemans C., Preneel B., Vandewalle J., and Dumortier J., *Revocable anonymous access to the Internet*, 2002, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.92.3215&rep=rep1&type=pdf

Goemans C. and Dumortier J. *Revocable anonymous access to the Internet*, 2003. https://www.law.kuleuven.be/icri/publications/440Retention_of_traffic_data_Dumortier_Goemans.pdf?where=

Kosta, E. and Valcke, P., 2006. *Retaining the Data Retention Directive*, Computer Law & Security Report, 22, 370-380.

McIntyre, T. J., 2008. *Data Retention in Ireland: Privacy, Policy and Proportionality*. Computer Law & Security Report, 24, 326-334.

Morariu, M., 2009. *How Secure is to Remain Private? On the Controversies of the European Data Retention Directive*, Amsterdam Social Science, Vol. 1(2): p. 46-65

Nabeth, T. and Hildebrandt, M., *FIDIS - Identity of Identity - Inventory of topics and clusters*, 2005. http://www.fidis.net/fileadmin/fidis/deliverables/fidis-wp2-del2.1_Inventory_of_topics_and_clusters.pdf

Renaudière P., *Perspectives on Data Notification, Data Retention and Data Protection from the European Commission*, http://www.lsec.be/upload_directories/documents/100209_DataProtection/3_Renaudiere_Philippe_EC_PerspectivesonDataProtection_1002009.pdf

Rowland D., *Anonymity, Privacy and Cyberspace*, 2000. http://www.bileta.ac.uk/Document%20Library/1/Anonymity,%20Privacy%20and%20Cyberspace.pdf

Taylor, M., 2006. *The EU Data Retention Directive.* Computer Law & Security Report, 22, 309-312.

Working Party on the Protection of Individuals with regard to the Processing of Personal Data, 1997. *Recommendation 3/97 - Anonymity on the Internet* (WP6)

Working Party on the Protection of Individuals with regard to the Processing of Personal Data, 1999. *Recommendation 3/99 on the preservation of traffic data by Internet Service Providers for law enforcement purposes* (WP25)

Working Party on the Protection of Individuals with regard to the Processing of Personal Data, 2000. *Working Document - Privacy on the Internet - An integrated EU Approach to On-line Data Protection* (WP37)

Working Party on the Protection of Individuals with regard to the Processing of Personal Data, 2005. *Opinion 4/2005 on the Proposal for a Directive of the European Parliament and of the Council on the Retention of Data Processed in Connection with the Provision of Public Electronic Communication Services and Amending Directive 2002/58/EC (COM(2005)438 final of 21.09.2005)* (WP113)

Working Party on the Protection of Individuals with regard to the Processing of Personal Data, 2006. *Opinion 1/2006 on the application of EU data protection rules to internal whistleblowing schemes in the fields of accounting, internal accounting controls, auditing matters, fight against bribery, banking and financial crime* (WP117)

Working Party on the Protection of Individuals with regard to the Processing of Personal Data, 2006. *Opinion 3/2006 on the Directive 2006/24/EC of the European Parliament and of the Council on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC* (WP119)

Working Party on the Protection of Individuals with regard to the Processing of Personal Data, 2008. *Opinion 1/2008 on data protection issues related to search engines* (WP148)

UNIVERSAL SERVICE
Up To Date Content Development Debate

Francesca de Gennaro
Autorità per le garanzie nelle comunicazioni (Agcom), Centro Direzionale, Isola B5, Napoli, Italy
f.degennaro@agcom.it

Keywords: universal service; scope of universal service; minimum set of services; provision of access at a fixed location; telephone directories; directory enquiry services; public pay telephones; low incomes; disable users; affordable price; quality of service; Internet access; broadband access; EU telecoms reform; being on line; bandwidth; competition to broadband widespread.

Abstract: The aim of the paper is to provide an overview on the universal service content development.
The paper introduces the universal service scope and the “universal” meaning regarding the communication sector together with the definition given by the European and the Italian framework as a “…*minimum set of services of specified quality to which all end-users have access, regardless of their geographical location and offered in the light of specific national conditions, at an affordable price*”.
After having described the universal service key elements, the paper drives through the recent content development in line with market needs, such as directory enquiry services, public pay telephones, internet access.
Then the paper focuses on the up to date broadband access debate. In particular, it is discussed if “being on line” has to be considered as a citizen right to the proper self fulfillment in the society or not.
In conclusion, the paper challenges to evaluate if the universal service principles has to contribute in achieving the broadband widespread or it’s up to the competition leading the required coverage.

1 SCOPE OF UNIVERSAL SERVICE AND “UNIVERSAL” MEANING

What is universal service?

It’s a legal instrument which aims at guaranteeing general interest services, in a liberalised environment.

The process of liberalisation of the telecom sector in the late 1990s went together with universal service rules. More in detail, whereas the market alone did not deliver basic and essential communication services and therefore needs of all the end-users were not adequately met, regulations on universal service have been introduced.

The aim of the universality is to prevent social exclusion. It is to assure that the provision of a defined “minimum set of services” of a specified quality is available to all EU citizens, irrespective of their geographical location and, in the light of specific national conditions, at an affordable price.

2 UNIVERSAL SERVICE DEFINITION

What does it mean a “minimum set of services”?

The Universal Service Directive, 2002/22/EC, article 1, paragraph 2, defines universal service as a *“… minimum set of services of specified quality to which all end-users have access, at an affordable price, in the light of specific national conditions, without distorting competition”*.

In line with the European regulatory framework, this concept is transposed into the Electronic Communication Code, article 1, paragraph 1, letter ll), where the universal service is illustrated as *“a minimum set of services of specified quality to which all end-users have access, regardless of their geographical location and offered in the light of specific national conditions, at an affordable price”*.

3 Universal service key elements

3.1 EUROPEAN AND THE ITALIAN FRAMEWORK: SERVICES FALLING UNDER THE UNIVERSAL SERVICE CONTENT

What does universal service cover?

According to the current European framework it covers the connection at a fixed location to the public telephone network and the access to publicly available telephone services at a fixed location (art. 4, Universal Service Directive).

More in detail, the connection made available shall be capable of allowing end-users to make and receive local, national and international telephone calls, fax and data communications, at data rates that are sufficient to permit functional Internet access, also taking into account prevailing technologies used by the majority of subscribers and technological feasibility.

In addition to the provision of access at a fixed location the Universal Service Directive incorporates in the universal service concept a number of other services that are closely associated with basic telephony, as they are necessary for consumers to make use of publicly available telephone services.

For these other services, however, the Directive leaves up to the Member States the decision to impose the obligation to provide them.

The mentioned additional services are the following: provision of directories and directory enquiry services (art. 5, Universal Service Directive), public pay telephones (art. 6 of the mentioned Directive), special measures for disable users (art. 7, Universal Service Directive) and tariff options or packages to ensure low incomes or consumers with special social needs (art. 9, par. 2, Universal Service Directive) are not prevented from accessing or using the publicly available telephone services.

As to the Italian framework, the Electronic Communication Code refers to the universal service content (art. 54, 55, 56, 57 and 59, paragraph 2) as the provision of access at a fixed location, of telephone directories, of access to public payphones - including free access to emergency numbers, together with special measures for disable and low income users.

3.2 MAIN FEATURES: QUALITY AND AFFORDABILITY

As already pointed out the mentioned “minimum set of services” have to be made available to EU end users, independently of their geographic location, at a specified quality and at an affordable price thus the achieved quality and the reasonable charge have to be strictly monitored in the contest of universal service obligations.

To this aim it’s requested that operators designated to provide universal service publish adequate and up-to-date information concerning their performance, in accordance with the quality of service parameters, definitions and measurement methods set out. In particular, the data delivering performance targets have to be of a specified form and manner in order to ensure that consumers have access to comprehensive, comparable and user-friendly information. In this way monitoring compliance with these targets is also easily ensured.

Additional quality of service standards can be introduced to assess the results of undertakings in the provision of services to disabled end-users and disabled consumers.

As anticipated the provision of the universal service has to be affordable to consumers.

More in detail, affordable price means that tariffs have to be defined considering specific local conditions therefore in line with national prices and income. As a consequence national regulatory authorities shall monitor the evolution and the retail tariffs level of the identified services falling under the universal service obligation.

It’s foreseen further the introduction of tariff options or packages which depart from those provided under normal commercial conditions in order to ensure that low incomes or special needs users are not prevented from accessing or using publicly available telephone services. The conditions of these special offers have to be fully transparent, published and applied in accordance with the non-discrimination principle.

4 Universal service recent content development

4.1 DIRECTORY ENQUIRY SERVICES, PUBLIC PAY TELEPHONES, INTERNET ACCESS

In compliance with the European framework, it’s possible that the concept of universal service evolves to reflect advances in technology, market developments and changes in users demand.

As communication market continues to progress in the light of the recent technological developments and competitive market conditions, the scope of universal service is currently under discussion.

More in detail, Europe is now focusing on verifying whether rules and definitions on universal service need to be updated. As the current concept of universal service was designed for basic communication services it’s wondering if is this approach is still valid today.

In particular the issue is related to the “minimum set of services” that, as defined in the Universal Service Directive, guarantees that people living in rural and remote areas or on a low income have affordable access to essential communication services and whether these services are still tailored to market needs.

Among the identified “minimum set of service” that could experience the universal service content development fall, at the moment, directory enquiry services, public payphones and Internet access.

At the present time, should they be included or excluded from the scope of the universal service?

For what concern directory enquiry services, with reference to the Italian telecom market, because of numerous existing offers, in terms of availability, quality and affordable price, the provision of directory enquiry services is now open to competition. As the market can deliver it, this service is not included in the universal service obligation.

As mentioned, the present discussion regards the provision of public pay telephone as well.

Owing to the development of mobile services and mobile coverage and to the end users habits, it’s evident a substitution effect of the public pay telephone with the mobile telephony.

This trend is leading to reconsider whether this provision is still included in the “minimum set of services” defined in the Universal Service Directive.

The Italian telecom market, it’s now experiencing a consistent reduction of traffic originating from public pay telephone in terms of minutes of conversation. In the light of the relevant change in consumers habits, Italian regulation has recently reviewed public pay telephone distribution criteria foreseeing a decrease of the number of public pay phone all over the territory. At the same time it has being ensured the availability of this service in specific location so to guarantee relevant social needs, therefore maintaining this provision in the scope of universal service.

With reference to the Internet access, the profitable debate is going on at European level is focusing on high-speed Internet connection for all citizens.

As already underlined a fundamental requirement of universal service is to provide users on request with a connection to the public telephone network at a fixed location. Now the requirement is limited to a single narrowband network connection, but once recognised the economic growth is linked to the widespread deployment of broadband services could be the scope of the universal service extended to cover broadband access?

5 Broadband access debate

5.1 “ACCELERATING BROADBAND ACCESS FOR ALL EUROPEANS”

There is a general tendency of European national governments to promote ad support end users universal access to and use of services and applications through electronic communication networks.

The rising Internet spread is giving a great contribute to the decrease of Digital Divide, considered as not access to the technologies allowing and guaranteeing information. This is preparatory to the Knowledge Divide reduction, that is to say obstacle to the learning.

The freedom of access to new technologies represents a challenge of social development and of democracy. Supporting the Internet access growth means to encourage the right to use information, it means to drive to take the opportunities offered by the modern technological age we are living in.

As Internet contents require today specialized networks of a determined bandwidth letting high rate of data transmission and assuring a definite quality of service, attention is focusing on speeding up Internet broadband access.

In line with this trend, EU telecoms reform package approved in November 2009, would lead, among others, to strength “*competition and consumer rights on Europe’s telecom markets, facilitate high speed Internet broadband connections to all Europeans*”.

According to the European Commission “*high-speed Internet is the passport to the Information Society and an essential condition for economic growth. This is why it is this Commission's policy to make broadband Internet for all. We need to combine all efforts to make sure that all citizens can get connected soon*”.

More in detail, one of the twelve top reform the European institution have settled on deals with “*accelerating broadband access for all Europeans*”.

The reform aim at overcoming the existing digital divide, as in rural areas there is still part of EU population, about 30%, that cannot access to broadband network connection, by better managing radio spectrum and by making it successfully available for wireless broadband services in regions where building a new fibre infrastructure results still too expensive.

Guarantee the Internet access facilitating the deployment of broadband technologies is a great opportunity to bridging the broadband gap considered not only in terms of access but also of usage and cultural divide.

As a result, could the endorsement of high speed connections all over Europe lead to recognise the broadband access as a fundamental right to be ensured thought the universal service scope to all citizens?

Moreover could “being on line” be correctly considered as a citizen right to the proper self fulfilment in the contemporary society?

The need to support the expansion of broadband connection is recognised all over the world.

One up to date example is given by South Africa commitment. The World Championship revenues South Africa got represent a powerful resource to accelerate investments in fibre optic infrastructure.

Internet traffic increase and development on one side and a great event as the World Championship on the other side will represent an explosive mixture to encourage growth. Successful broadband infrastructure deployment has been a tough goal for South African government that is now carrying out extensive collaboration and partnership among private companies and industry.

Sharing of fibre optic infrastructure will contribute to increase competition and, at the same time, to give citizens access to high speed connection at lower prices following the slogan '*more Internet, more connection to the rest of world*'.

5.2 A DEFINITE BANDWIDTH IN THE SCOPE OF UNIVERSAL SERVICE

With the increase of broadband connection and the significant economic impact and benefits of broadband Internet, Europe is now wondering whereas a definite bandwidth is to be considered in the light of universal service.

As broadband has being recognised as a factor contributing to social and economic development, capable to stimulate economic growth also overcoming the cultural divide, is it to fall under the scope of universal service?

Within the telecom reform, European institutions handled adjustments to the fundamental principles of the regulatory framework for electronic communications network and services.

More in detail, in the contest of the review of the EU regulatory framework for e- communications dealing with the universal service content, a recital – recital 5 – has been introduced in the Citizens' Rights Directive, 2009/136/EC, which complements the recital 8 of the Universal Service Directive, 2002/22/EC.

The recital 8 of current regulation on the topic of universal service requirement regarding functional Internet access limits "*to single narrowband connection*" and to "*currently available voice band modems that typically offer 56 kbit/s*" specifying that "*connections to the public telephone network at a fixed location should be capable of supporting speech and data communications at rates sufficient for access to online services such as those provided via the public Internet*".

The new recital introduced in the Citizens' Rights Directive, recital 5, dealing with functional Internet access while always referring to the consideration that data connection, in the light of universal service, should be able of allowing data communication at rates adequate to permit functional Internet access suggests to take "*due account of specific circumstances in national markets, for instance the prevailing bandwidth used by the majority of subscribers in that Member State, and technological feasibility, provided that these measures seek to minimise market distortion*".

Could the "*prevailing bandwidth used by the majority of subscribers*" lead to focus on the universal availability of broadband connection, at an affordable price?

6 Universal service or competition to broadband widespread?

As previously pointed out the Universal Service Directive disposed that universal service content will be revisited more than once due to technological developments and competitive market conditions.

As a result, following up the significant social and economic developments occurred in the telecom sector as illustrated above, has the universal service scope to be reviewed taking account of the broadband growth or is the competition so strong and fruitful to deliver broadband network connection ensuring social inclusiveness?

Should universal service principles help to reach EU goal of "*broadband for all*" or would competition on the open telecom market be more effective?

For the time being it seems that the market alone cannot yet be relied on to meet the demand for broadband connection services in market failure areas, as well.

Definitely it can be asserted the universal service is still playing an important role in ensuring all citizens social inclusion.

REFERENCES

Universal Service Directive 2002/22/EC.

Electronic Communication Code (Legislative decree n. 259/2003).

Citizens' Rights Directive 2009/136/EC.

European Commission, IP/10/218, *Telecoms: consultation on future universal service in digital era.*

European Commission, *Consultation on universal service principles in e-communications.*

European Commission, MEMO/09/513, *EU Telecoms Reform: 12 reforms to pave way for stronger consumer rights, an open internet, a single European telecoms market and high-speed internet connections for all citizens.*

European Commission, IP/08/1397, *Broadband Internet for all Europeans: Commission launches debate on future of universal service.*

Dods D., Brisby P. et al., *Reform of European Electronic Communication Law: A Special Briefing on the Radical Changes of 2009*. CTRL, The Journal of E-Commerce, Technology and Communications, issue 4 2010.

Carli S., *L'Africa vola sulla fibra ottica con i soldi dei mondiali di calci*. AFFARI & FINANZA. 7 June 2010.

THE DILEMMA OF DIVIDING THE DIGITAL DIVIDEND

Manuel J. Fernández Iglesias, Luis M. Álvarez Sabucedo
ETSE de Telecomunicación. University of Vigo. Spain
manolo@uvigo.es, luis.sabucedo@uvigo.es

Keywords: digital dividend, analogue switch-off, spectrum assignment, cognitive radio, frequency planning.

Abstract: Europe is involved in a process of transition to digital terrestrial television aimed to replace all analogue broadcasting infrastructures for digital ones by year 2012. Due to a better efficiency on the use of the transmission spectrum, the analogue switch-off will liberate some of the channels previously allocated to analogue television, a resource that will be available for new telecommunication services. Public Administrations at the European and national level will play a leading role in deciding to which services this 'digital dividend' will be granted and at which costs. A convenient return for the citizenship, both financial and social has to be guaranteed. This is the dilemma of dividing the digital dividend.

1 INTRODUCTION

Both traditional analogue and new digital television stations use the UHF band for broadcasting of television programs. The UHF or Ultra High Frequency band ranges from 300 megahertz to 3 gigahertz. In addition to television, this part of the electromagnetic spectrum hosts services like aviation and maritime radio, mobile telephony, amateur radio, satellite radio, and even is the preferred band for garages' remotes or the microwave ovens we have at home. In fact, this is a hugely saturated band.

You could say that most of the waves that surround us are in the UHF range. Even in the small slot devoted to television, more or less the 15% of the UHF spectrum, the saturation at this time prevents in most areas in developed countries the planning of new television programs, as channels are fully occupied by other television stations, both analogue and digital.

The good news is that technological advances in the field of digital communications are enabling a more productive use of that part of the spectrum, as digital protocols and optimization techniques make more efficient use of it. We may compare the present use of analogue broadcast radio channels with the use of a highway that would make big sport utility vehicles (SUV) occupied only by their drivers. They would use most available space, it would be difficult to share the highway with them for fear of colliding, and probably these drivers would reach the end of their journey in a worse condition than that of departure. However, in digital, the situation would resemble the distribution of travellers in comfortable buses driven by cautious trained drivers. Many buses would occupy the highway without fear of collision, and each bus would host a lot of people. And upon arrival, passengers would be in the same condition in which they were when departed.

As a result of the efficiency of digital broadcast, and despite the increased number of programs available to the viewer, a significant part of the frequencies used by plain old TV services will be released when the deployment of digital terrestrial television (DTT) is completed and the associated analogue switch-off is performed in 2012. Those liberated frequencies or channels are named the 'digital dividend'. In other words, with the demise of analogue television, a series of UHF channels previously allocated to it will be released, channels that will be available to accommodate new telecommunications services.

The UHF band is regarded as high quality 'prime spectrum' because it combines range with a great ability to transmit signals. For these features, the spectrum released is being claimed by telecommunication service providers and broadband operators, mobile operators, public administrations for official and emergency communications, and even existing television stations to expand their present-day program offer.

The use of the spectrum released by analogue TV could facilitate the provision of these services at very attractive costs, because they would require to deploy a number of transmission facilities far smaller than in the case of higher frequencies.

In the management of this new situation, a number of new approaches and technologies are under study. In particular, we outline below a new technology, currently under development, that can play a paramount role in this context. One problem to be faced is related to properly dealing with allocation gaps in the spectrum and efficiently utilizing them in different

circumstances. In this context, the introduction of the so-called cognitive radio, a novel approach that enables a smart use of bandwidth by monitoring the actual net load at any time, may increase the efficiency of the entire system.
This situation poses significant challenges to public administrations, as they are responsible for managing and allocating these resources to operators, trying to maximize at the same time the benefits to the citizen and the social return of the new situation. In the case of Europe, a window of opportunity appears for public administrations to make the best possible use of this scarce resource, taking into account present and future citizen needs. The digital dividend could be used, for example, to facilitate access to broadband communication in isolated or less advanced areas, and thus help to overcome the digital divide. It could also support new mobile services ranging from eGovernment to new communication solutions for emergency situations.
The rest of the paper is devoted to analyse the concept of digital dividend according to several relevant stakeholders, defining characteristics as its value, the type of services it may enable, and the forces that drive an eventual decision on its allocation. Then, we introduce cognitive radio as an alternate approach that is presumed to have an important role on optimizing the use of the spectrum, and on the provision of non-licensed application and services. We will discover how cognitive radio may help regulators and public administrations to fill in the gaps when allocating spectrum in the future. Next, we offer some insight on how the aspects discussed so far are being put into practice in a study case, namely the allocation process consequence of the analogue switch-off in Spain. Finally, we present some conclusions and propose a roadmap for public administrations to manage the digital dividend.

2 The Digital Dividend

We strongly encourage authors to use this document for the preparation of proceedings. Please follow the instructions closely in order to make the volume look as uniform as possible. (Moore and Lopes, 1999).

At the present time, terrestrial TV is supported by the transmission of analogue signals in most parts of the world. Back in mid 20th century, broadcasting was one of the first major telecommunication services requiring planned use of the public spectrum. According to the state of the art at that time, the UHF band was selected as the optimal solution in terms of propagation and coverage. This range of frequencies offers an unbeatable balance of capacity and range. State-of-the art technology guarantees coverage of large geographical areas with relatively few base stations. As a consequence, network roll-out can be achieved at lower costs when compared to services delivered at higher frequencies, which offer greater bandwidth but at shorter range (see Figure 1).

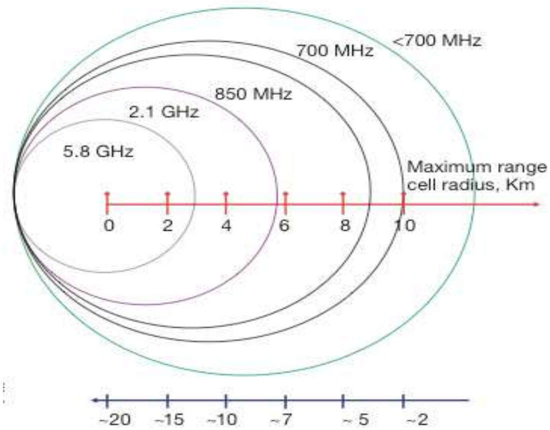


Figure 17. Number of base station and its size for each frequency considered. Source: BBC R&D

The band of operation also influences the cost of user terminals. For example, antennas are simpler to design at lower frequencies. This way it will possible to produce cheaper devices, along with the related infrastructure, making it possible for a larger number of people getting into the newly provided solutions and, thus, increasing the potential market.

After the analogue switch-off, stakeholders will rush for their share of a resource – the UHF band – that generates around €800bn in net present value for the European economy. On the other side, spectrum is a national resource that should be carefully planned and optimally allocated to both commercial and public services.
Two questions arise: everyone will get its share? How large is the digital dividend? Indeed there are many candidate services that claim their share of the digital dividend, the most active being TV stations and mobile operators. All of them offer their own perspective of the issue and, of course, they are willing to participate in the exploitation of this new resource.
According to the European Broadcasting Union (EBU, 2008), broadcasting services are called to make a large usage of this recently freed bandwidth, at least in a significant part. Nevertheless, practitioners must be aware of the differences in terms of actual coverage and legal framework existing on each country and, therefore, particular considerations that must fulfilled in each case. Depending of particular constraints, EBU suggests to reserve channels from 21 to 55 for television broadcast, that is, 200 Mhz distributed in 25 8-Mhz channels.
EBU also recognizes that other stakeholders may enter this market, but they point out that to share this spectrum with non-broadcasting services, either fixed or mobile, further research is needed to guarantee the technical viability of newly proposed solutions. EBU fears that a suboptimal approach would lead us to waste too much space on guard bands to overcome propagation problems, which provide little return of investments. Also, during this planning phase, it should be beard in mind a large range of different options that should include some non-

commercial usages such as Services Ancillary to Broadcasting or Services Ancillary to Programme (Doeve, 2007).
On the other side, according to (Forge et al. 2008), from an economic point of view, this resource should not be fully assigned to broadcasting services as the return from mobile services is bigger. From this study, it is estimated that the economic output per MHz of bandwidth is €168 million for mobile (non-broadcasting services) compared to €28 million for the digital TV case. Actually, this report is also prone to the investment in mobile services due to other factors such as the larger number of jobs created and bigger indirect economic effects.
Also, it should be considered some specific opportunities and challenges to be address in this context. Fist, the role of the digital dividend on closing the digital gap. This spectrum, due to its highly favourable propagation conditions, could be the perfect option to provide with broadband access to people far from the areas with good network coverage at this time. Second, the chance of dramatically improving the support for health and elderly care applications. Due to the ageing population and the high cost of in-hospital care, the EU society must search for alternatives. In this line, an alternative would be the deployment of in-home care services by means of remote monitoring and assistance. For this task, this newly available bandwidth can be very important. Third, the digital dividend combined with techniques like cognitive radio, may support a flexible disaster recovery network for global warning. There is an opportunity to reserve part of this band to be able to send information to the population in case of natural or man-produced disasters or accidents. It is difficult to assess the economic value of this approach, but its social return cannot be denied.
To sum up, broadcasters see the window of opportunity, express their willingness to make the most of this free area of the spectrum, and implicitly state their (historical) 'property' rights. However, it is important to have an idea about the most likely future use for this resource to be able to design a rational use plan. From (Spectrum Value, 2008) we can infer the most relevant concerns from mobile operators. According to this report, data services, and more specifically email and browsing the web will be the more demanded services in the long term. Thus, the division of the spectrum must take this into account when planning non-broadcasting services. According to that report, mobile operators claim more than 92MHz to provide these services in nomadic scenarios, which corresponds to 25% of the UHF broadcast band and around 40% of the estimated digital dividend. Besides, decisions on this area should be carefully evaluated, as the dividend may generate extra incomes ranging from 63 billion to 165 billion Euro during the next 20 years. For example, the dividend may be used to consolidate the triple play market of mobile operators, offering mobile TV and nomadic data services at a convenient cost.
On the other side, governments see frequency channels as a finite national asset, they see spectrum as a national resource. In fact, UHF already hosts essential communication services in

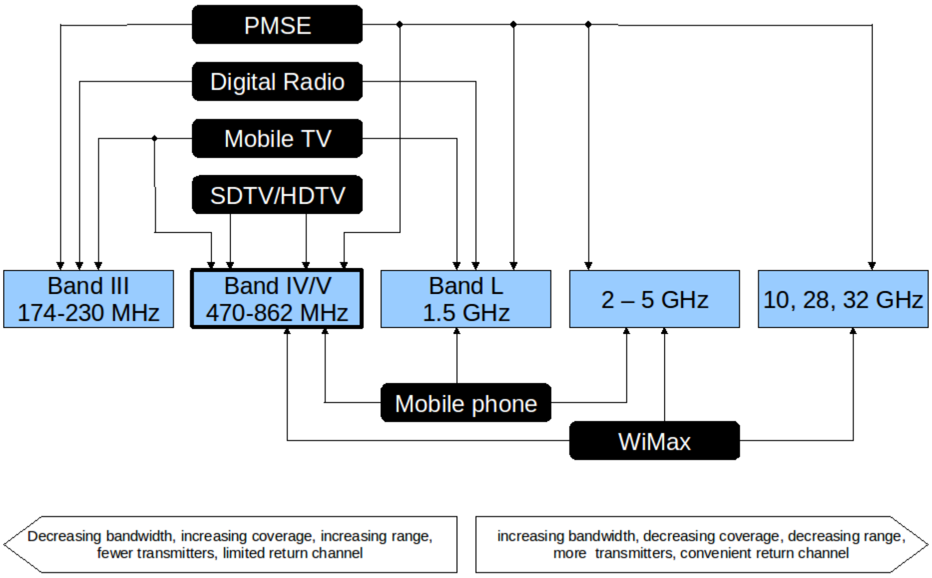


Figure 18. Alternative uses of frequency bands

defence, transport, emergency services and health care.

As a consequence, non-confluent driving forces influence the way administrations at different levels are facing this situation:

- Public administrations should guarantee the re-allocation of previously licensed private TV stations, in a way that their potential audience (i.e. potential income) is not compromised. This includes maintaining the agreements signed with operators to minimize network roll-out due to the digital upgrade, which in many cases included granting additional TV and multimedia services to existing analogue operators.
- Public administrations have a one-in-a-lifetime opportunity to deploy new public services as part of the return of the digital dividend. This includes public TV stations or programmes, fostering other services already operating in the UHF band, and providing new wireless services in remote areas to overcome the digital divide.
- Public administrations have to cope with market pressure, mainly from mobile operators, broadband operators and TV stations.

Before this resource is put at the disposal of practitioners to be split, a previous step should be undertaken. Not only broadcasting TV is using this range of frequencies at this time, but other services such as aeronautical radar or astronomic communications are long-term occupants of the UHF band.

At the present moment, it seems to be a general agreement about moving these services to another bands and configure larger chunks of available spectrum (c.f. Figure 2). In fact, most countries in Europe, with slight differences on the exact frequencies, have decided to free the 800MHz band. There are several benefits from this decision:

- As this range of frequencies are in the lower part of the UHF spectrum, the cost of the equipment will be lower. Also, the more countries free this region of the spectrum, the more devices will be produced. Thus, due to economies of scale, again, the prices of the devices will be more competitive.
- As it is increased the available spectrum, it is expected to increase also the number of possible competitors in the market. This will benefit users in the long term, as offer will spread and promote competition among a larger number of players.

One of the pending issues of the 800MHz band clean-up is the reallocation of Programme Making and Special Events (PMSE) services. Migration of PSME providers should be facilitated by supporting the migration to new equipment, by providing the proper training, and by developing a safe and clear regulation to protect their services.

However, the clearance of the 800 MHz band may be too expensive or too complex. The benefits estimated, just in UK and according to a conservative approach, are about £2-3bn in net present value (Forge et al., 2008). On the other side, in addition of the cost of new communication equipment, there is a hidden cost on this operation that must be taken into account related to the lack of international agreements on new uses of the 800 Mhz band. The actual definition of the spectrum limits, or even its purpose, are not exactly the same in all countries. These differences may lead to compatibility problems among different regions.

In addition to this, initiatives are being put forward to efficiently manage the interleaved spectrum, that is, parts of the spectrum not being used for planned services in a particular location that will be available for other services in a shared basis.

3 THE TIME OF COGNITIVE RADIO

The approaches to the use of the digital dividend discussed so far are characterized by a fixed spectrum, channel-based assignment policy. However, the inclusion of non-broadcast services in the UHF band plus the presence of services like PSME implies that a significant part of the spectrum will be used according to the patterns of mobile and data services, that is, it will be used sporadically with relevant geographical variations due to propagation and channel planning. In that cases, the utilization of assigned spectrum ranges from 15% to 85% with a high variance in time (FCC, 2003). Cognitive radio technology (Akyildiz et al., 2006; Mitola and Maguire, 1999) is expected to play a relevant role in exploiting wireless spectrum, overcoming inefficiencies in spectrum usage by detecting unused spectrum and sharing the spectrum without harmful interference with other users. Cognitive radio systems capture the best available spectrum according to users' communication requirements, and provide spectrum mobility by maintaining the quality of service during transitions to better spectrum when appropriate.

UK's agency Ofcom is pioneering the regulation of this new approach to spectrum management (Ofcom, 2009) by considering the use of interleaved spectrum by licence-exempt applications. For this, three approaches are under study to determining unused spectrum slots, namely: detection, geolocation and beacons (Fette, 2006). After a stakeholders' consultation process, some conclusions arise:

- beacons were identified as the least appropriate,
- geolocation was recognized as the most important mechanism in the short to medium term,
- detection was identified to require further research and development efforts,
- and detection-only devices being many years away according to the present state-of-the art in the field.

As a consequence, Ofcom asked stakeholders to further develop the concepts and algorithms for geolocation, and plans to launch a new consultation process later this year. In the United States of America, the Federal Communications Commission and the National Telecommunications and Information Administration are also working on the regulation of the use of cognitive radio and dynamic spectrum sharing, focusing on unlicensed devices below 900 MHz and in the 3 GHz Band (FCC, 2002; FCC, 2004).

Cognitive radio enables conceptual frequency allocation, that is, fixed frequency slots are not assigned to operators in a per-service basis, but licenses are granted provided the new services behave according to the cognitive radio model. For example, regulators may grant radio services provided these new services do not interfere existing licensed services, and they accept interference from any existing licensed service. IEEE is working on a new member of the 802.XX family of network access protocols to become the first standard based on cognitive radio (Cordeiro et al., 2005) . The charter of IEEE 802.22, the Working Group on Wireless Regional Area Networks (WRAN), is to develop a standard for a cognitive radio-based physical to medium access control to air interface for use by license-exempt devices on a non-interfering basis in spectrum that is allocated to the TV broadcast service.

4 THE SPANISH SITUATION AS A CASE STUDY

Like in most countries in Europe, UHF analogue TV broadcast allocation in Spain ranges from 470 Mhz to 862 MHz corresponding to Band IV/V channels 21 to 69 (CNAF, 2007). This makes 49 8-Mhz channels, which are fully utilized by licensed broadcasters (state, regional, local), by the TV transport network and PSME services, by non-licensed broadcasters that are operating due to specificities of the Spanish legislation outside the scope of this paper and, in some areas, by guard bands and trans-modulation equipment to avoid interferences among channels.

TV broadcasting takes 15% of the UHF band, a situation similar to that in most countries in Europe. The rest of the UHF band (300 Mhz to 3,000 Mhz) is utilized by personal radio services like FRS or GMRS, mobile phones, cordless phones, public service agencies, maritime and aeronautical communications, broadband data (e.g. WiFi), short-range multimedia (e.g. Bluetooth), radio-frequency RFID and UHRFID tags, amateur radio, satellite radio, remote command devices, and microwave ovens. Most of these services may claim its share of the digital dividend.

Once the analogue switch-off is completed in Spain by April, 2010, the Spanish government will assign a complete multiplex (i.e. a complete 8 Mhz UHF channel) to each of the 6 private TV stations operating in the country, plus 2 additional channels to the public operator TVE. In other words, nationwide DTT operators will consume 8 channels previously allocated to analogue television. Additionally, local and regional operators will utilize a variable number of channels depending on the region, this number ranging from 3 to 5. This makes between 13 and 15 8-Mhz channels allocated to DTT in a given Spanish (UHF-planning) region. Besides, 12 additional channels are estimated to be needed due to radio-electrical planning needs to guarantee the reception of all licensed TV services (guard bands, channels needed to solve interferences due to out-of-guard reception in single frequency networks, gap-filler planning, etc.)

To sum up, in the area of operation of 11 - 13 DTT multiplex-wide stations (i.e. 44 to 52 distinct TV programs) a total of around 200 Mhz will be needed only for DTT services. That makes 51 % of the former analogue TV UHF band, and complies to the EBU recommendations discussed above.

The situation described above liberates 192 Mhz of spectrum previously allocated to analogue TV, that is, 49 % of the former analogue TV band or 7.1% of the complete UHF band. As an example, this spectrum can be configured as 24 8-Mhz TV channels or 19 10-Mhz mobile telephone channels contributing to the digital dividend in Spain.

Spain is also working towards the liberation of the 800 Mhz band. The Spanish government plans to assign the upper 5 UHF channels (Band IV/V UHF channels 65 to 69) to mobile phone operators after the switch off, a total of 40 Mhz that represent a 10.2 % of UHF band IV/V, and 20.8 % of the digital dividend in Spain. Note that this figure is far below the spectrum recommended for mobile operators by Spectrum Value Partners (2008). For this to be completed, it is needed to reallocate digital TV services presently using these channels to lower channels in the band, something that will require relevant investments from TV stations.

No decision has been taken so far on other uses discussed in this paper. The Spanish regulations on the use of the spectrum (Spanish Government, 2008) establishes the foundations that will drive spectrum assignment in the future. These regulations establish that frequency planning and assignment should be driven by the principles of efficiency and effectiveness. However, frequency assignment is based on a planning process that bind frequencies and channels to specific services, as defined by the National Chart of Frequency Assignments and the technical plans for television and radio.

On the other side, the Spanish telecommunications law (Spanish Government, 2003) created the National Radiocommunications Agency as a convergent regulation entity centralizing all regulation initiatives presently scattered in several departments. This body, similar to UK's Opcom, will guarantee a rapid and flexible response to the new challenges in spectrum planning, including the management of the digital dividend. Regretfully, this body has not been deployed yet.

5 CONCLUSION

As a consequence of this situation, we propose a roadmap for public administrations to guarantee the highest possible return for the citizen from the digital dividend (Fernandez-Ribas, 2009). This includes finding the right balance between income obtained for licensing spectrum, and the new public services offered or existing public services improved due to the part of the digital dividend reserved for them. This roadmap is defined as follows:

- Refarm the UHF band to obtain the largest chunks of free spectrum possible.
- Define the amount of the dividend that will be put into market, and the part that will be managed by the administration to offer new public services or improve existing ones.
- Select the services that will be favoured in both cases. Analyse the convenience of putting into market service-neutral chunks of spectrum, either in a fixed-channel model or according to the cognitive radio model.
- Select a licensing method (e.g., auctioning, secondary frequency market, public call for tenders, taxes, etc.)

The last three steps may be performed iteratively until convergence to a satisfactory solution is reached. For this, feedback from market forces, social forces, and the experience or recommendations from other administrations is needed. Once the right balance is found, it is time to proceed. However, experience shows that operators claim for rapid decision procedures, and licensing periods as extended in time as possible to guarantee investment returns. As a consequence, market pressure will require from regulators a long-term decision as rapidly as possible. Market-driven licensing methods like auctioning or the configuration of a secondary frequency market may seem very attractive for governments because of their flexibility, the reduction of management overheads, and the promise of higher revenues. However, according to Bohlin, Weber and Preissl (2006), these methods may generate problems that do not compensate the higher economical benefits expected. Secondary markets may increase costs generating a feeling of frequency shortage if spectrum is monopolized by global-operating, for-profit organizations. Besides, a radical liberalization may generate market fragmentation and technical problems for secondary intensive non-exclusive uses. On the other side, auctioning may prevent non-for-profit uses of the spectrum.

As a consequence, public administrations should manage radio-electrical spectrum as a public resource, selecting the most appropriate mix of licensing method to guarantee maximal social return.

With respect to the services to be provided, apart from additional TV channels and mobile services, other approaches should be considered. As pointed out, some see this situation, especially in public administrations, as a one-in-a-lifetime opportunity to fight back the digital divide. In areas where population dispersion and a difficult orography are the driving aspects insofar telecommunication infrastructure deployment is concerned, the digital dividend may be used to provide broadband access to settlements where typical wired options are too costly. For this, the European Commission promotes the identification of a reserved harmonized sub-band of the UHF band for bi-directional broadband communications, and favours a flexible channel plan with different options to be applied in particular national situations (EU, 2007). This initiative has the support of the European Telecommunications Network Operators' Association (ETNO, 2008).

Collectives presently refrained from accessing the net may utilize the upper part of the liberated channels to be connected. As mentioned above, the frequencies liberated as a consequence of the analogue switch-off have near-optimal propagation and attenuation characteristics. Applying wireless technologies (Labiod, 2007) like

Wifi (Arbaugh, 2001) or WiMax (R. El-Azouzi, 2008), access to the Internet, that is, the right to access to a basic service for the citizen (Reding, 2007), can be guaranteed. An issue that must be coped with is related to the efficient use of the spectrum. As this is a limited and valuable resource it must be used wisely. In order to make it possible, different approaches must be taken into account:

- Bandwidth usage must be conveniently planned. This includes, as mentioned above, different strategies such as the clearance of marginal bands assigned to services possible on other suboptimal bands.
- Once the channels are set, suitable technologies must be chosen in order to guarantee the best possible use for the resource. In this line, authors suggest the use of cognitive techniques as they enable services in small geographical areas with little or no interference with other existing services.
- Economical revenues must be seriously considered. According to the referenced literature, differences may be very relevant from a portfolio of services to a different one. Also, the digital dividend may be used to undertake social issues and face situations as the digital divide or in-house care for elderly people.
- The role to be played by the regulation entities from different countries must not be underestimated, specially in the European case, where countries have efficient coordination tools, but are prone to local pressures from stakeholders, arguing that UHF line-of-sight ranges enable local planning. It is quite important to guarantee an homogeneous treatment of band usages to maximize the outcomes of this resource.

To sum up, this is a one-in-a-lifetime opportunity. Success or failure on managing the digital dividend will be with us for years to come, and from its accurate allocation a large number of benefits can be derived to the entire society.

ACKNOWLEDGEMENTS

We want to thank Ministerio de Educación y Ciencia for their partial support to this work under grants “Servicios adaptativos para e-learning basados en estándares” (TIN2007-68125-C02-02) and Xunta de Galicia, Consellería de Innovación e Industria through projects “SEGREL: Semántica para un eGov Reutilizable en Entornos Locais” (08SIN006322PR)

REFERENCES

Akyildiz, I.F., Lee, W-Y., Vuran, M.C. and Mohanty, S. (2006). 'NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey', Computer Networks 50(13), pp. 2127-2159

Arbaugh, A. W.; Shankar, N.; and Justin Wan, Y.C. (2001). 'Your 802.11 Wireless network has no clothes', IEEE Wireless Communications pp. 44—51

Bohlin, E., Weber, A., and Preissl, B. 'How free is the radio spectrum? '. Info - The journal of policy, regulation and strategy for telecommunications, 8(2), pp. 3-5.

Cordeiro, C. Challapali, K. Birru, D. Sai Shankar, N. (2005). 'IEEE 802.22: the first worldwide wireless standard based on cognitive radios', in Procs. of the 1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005. Baltimore, MD, USA. pp. 328-337

CNAF (2007). Cuadro Nacional de Atribución de Frecuencias. Notas de Utilización Nacional (CNAF-UN). Dirección Técnica de la Comisión Nacional de Atribución de Frecuencias. Spanish Ministry of Industry, Tourism and Commerce. 2007.

Doeven, J. (2007) 'Implementation of the Digital Dividend. Technical constraints to be taken into account'. EBU Technical Review, January 2007.

EBU (2008). European Broadcasting Union. EBU Comments to the RSPG. Opinion on EU spectrum policy implications of the digital dividend.

El-Azouzi, R. and Peyre, T. (2008). Capacity at the mac layer. Radio Resources Management in WiMAX from Theoretical Capacity to System Simulation, ISEP(9781848210691):15, September. 2008.

EU (2007). European Commission. Reaping the full benefits of the digital dividend in Europe: A common approach to the use of the spectrum released by the digital switchover. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions.

ETNO (2008). European Telecommunications Network Operators' Association. ETNO Reflection Document on the Digital Dividend in Europe.

FCC (2002). Additional Spectrum for Unlicensed Devices Below 900 MHz and in the 3 GHz Band. Docket No. 02-380.

FCC (2003). Notice of proposed rulemaking and order. Docket No. 03-222. December 2003.

FCC (2004). Unlicensed Operation in the TV Broadcast Bands. Docket No. 04-186

Fernandez-Ribas, A. and Shapira, P. (2009). 'The role of national and regional innovation programmes in stimulating international cooperation in innovation'. International Journal of Technology Management 48(4), pp. 473 - 498.

Fette, B.A. (2006). Cognitive Radio Technology. Newnes Communication Engineering series.

Forge, S., Blackman, C., and Bohlin, E. (2008) 'Economic Impact of Alternative Uses of the Digital Dividend'. Intereconomics. May/June 2008. pp. 149-162.

Labiod, H., Afifi, H., De Santis, C.. Wi-Fi, Bluetooth, Zigbee and WiMax. 2007, XVI, 316 p., Hardcover. ISBN:978-1-4020-5396-2. Springer.

Mitola, J., III and Maguire, G.Q., Jr. (1999). 'Cognitive radio: making software radios more personal,' Personal Communications 6(4), pp.13-18, IEEE.

Ofcom (2009). Digital dividend: cognitive access. Statement on licence-exempting cognitive devices using interleaved spectrum. Ofcom Statement. 1 July 2009.

Spectrum Value Partners (2008). Getting the most out the digital dividend. Allocating UHF spectrum to maximize the benefits for European society.

Reding, V. (2008). 'How Europe can Bridge the Broadband Gap'. Speech to the Bridging the Broadband Gap Conference. March 2008. European Commission.

Spanish Government (2003). Ley 32/2003, de 3 de Noviembre, General de Telecomunicaciones. Boletín oficial del Estado no. 264. 4th November, 2003.

Spanish Government (2008). Real Decreto 86372008, de 23 de mayo, que deroga la Orden de 9 de marzo de 2000 por la que se aprueba el reglamento de desarrollo de la Ley 11/1998 General de Telecomunicaciones, en lo relativo al uso del espectro radioeléctrico (modificada por el Real Decreto 424/2005. Boletín oficial del Estado no. 138. 7th June, 2008.

Consent, an instrument for patient empowerment?

Griet Verhenneman
ICRI, K.U.Leuven, St. Michielsstraat 6, b3443, Leuven, Belgium
griet.verhenneman@law.kuleuven.be

Keywords: Consent, medical treatment, data protection, patient empowerment.

Abstract: This paper deals with the use of consent in eHealth environments. The patient's consent is on the one hand required for treatment, and since the introduction of the Data Protection Directive on the other hand for data processing. Three factors make the use of patient consent in healthcare environments to an increasingly complicated issue, eventually risking to backfire the empowerment of the patient. Different requirements are first of all set to the two types of consent, the number of occasions in which the patient is requested to consent is secondly steeping fast and consent is thirdly becoming an instrument to express not only agreement, but also refusal. Consequently, chances are that healthcare practitioners and patients are confronted with a "consent-overload". In this changing environment the lifecycle management of the patient's consents becomes a true need. Can an advanced lifecycle management system "re-empower" the patient? In order to find an answer to that question this paper moves from a discussion of the legal theory on consent for medical treatment and data processing in Belgium to a more practical view on how to handle consents during a full lifecycle, presenting the results of the Belgian Share4Health project.

1. INTRODUCTION

To put it blunt, consenting in a healthcare environment used to imply nothing more than singing a piece of paper on the request of a doctor. Three trends now cause a change in this rather simple act of consenting. The healthcare environment too is firstly moving from a paper to a paperless environment. Secondly a change is noticeable in the attitude of patients towards healthcare practitioners. Not only have patients become a lot more articulate, the variety in available healthcare options is increasing steeply, often leaving the patient with a (difficult) choice he would not had worried about before. Last but not least, the change towards a computerized healthcare environment brings along an enormous increase in the amount of patient data being automatically processed, stored and exchanged, making consent a well used legal ground for data processing. And so the concept of consent is these days an omnipresent critical link in the healthcare environment. Consent as an instrument for patient empowerment, consent as a form of legal evidence and last but not least consent as a legal ground.

Two other factors complicate this critical link even further, eventually risking to backfire the empowerment of the patient: a) a different set of requirements needs to be taken into account when dealing with either consent for treatment or consent for data processing and b) since the introduction of electronic patient files and electronic sharing systems consent has become an instrument to also express refusal and not only agreement.

Consequently, chances are that healthcare practitioners and patients are confronted with a "consent-overload". Life-cycle management of the patient's consents then becomes a true need. Can an advanced life-cycle management solution safeguard the empowering function of patient consent?

2. LEGAL DISCUSSION

2.1 Patient consent for treatment and for data processing compared

Within a medical setting two types of consent are required for two different, but often overlapping actions: the act of treatment and the act of data processing. The requirement of an informed consent for medical interventions on the one hand is an application of the patient's rights of self determination and physical integrity as protected by not only several legal sources, but also ethical sources, including medical deontology[1]. The requirement to obtain an informed consent for data processing on the other hand follows from article 8 of the Data Protection Directive, marking the informed consent as a legal exemption for the processing of sensitive data.

Although both types of consent thus spring from different legal sources, there validity depends on the same principles: a consent needs to be given freely, before the action starts, after being informed and for a specific purpose. Legislation on both types of consent does however not always interpret or implement these principles in the same. A large difference exists moreover with regard to the required form.

2.1.1 Free

As a free consent can only be considered a voluntary decision, taken by an individual in possession of all of his faculties and in absence of coercion of any kind, be it social, financial, psychological or other. On this point medical regulations and data processing regulations agree.

A consent given – either for treatment or for data processing – under the threat of non-treatment or a lower quality of treatment cannot be considered free. A consent given by a data subject who has not had the opportunity to make a genuine choice or who has been presented with a *fait accompli* does not meet this first condition either.

With regard to the consent for data processing, the Article 29 Working Party furthermore confirmed that it is misleading for a doctor to legitimize the processing of health data through consent when the processing operation is a necessary and unavoidable consequence of the medical situation: "*Reliance on consent should be confined to cases where the individual data subject has a genuine free choice and is subsequently able to withdraw the consent without detriment*", the Working Party concluded.

[1] The principle has been adopted in the European Manual for medical ethics, the European Charter of the Rights of the patient, in the Declaration of Amsterdam concerning the promotion of the rights of the patient and the Treaty of human rights and biomedicine.

2.1.2 Specific

In general it can be said that a consent is considered 'specific' when it relates to a well-defined, concrete situation.

With regard to consent for medical treatment a separate consent is more concrete demanded for each specific intervention. Examination acts, treatments and decisions to continue or refuse such acts are all considered as separate interventions. In practice however, it is under discussion whether the healthcare provider has to obtain a separate consent for so called 'extended operations'. These are operations which are additional to the interventions originally foreseen and of which the necessity only raised during the execution of the intervention to which the patient had consented. Different lines of thinking exist in literature and jurisprudence, but the most honoured is the strict line of thinking. According to that strict line, a physician can only change the scheme of the intervention without the consent of the patient in case of an emergency. An extended operation can thus only be performed without an additional consent when the medical duty of the healthcare practitioner prevails over the consent requirement.

With regard to consent for data processing, a consent is regarded specific when it concerns the processing of a defined set of data, for a delineated purpose, by a determined individual or category of persons. Consequently, a 'general agreement' of the data subject e.g. to the collection of his medical data for an electronic health record and subsequent to transfers of these medical data of the past and of the future to health professionals involved in treatment, would not constitute a specific consent. It moreover follows from the requirement that consent is a one to one action and cannot be one to many. An example of an acceptable category of persons to whom a consent can be given is the treating medical team.

2.1.3 Informed

Under Data Protection Regulations, consent is not regarded valid unless it is given after being provided with the necessary information. By informing the patient, the latter needs to be able to counterbalance the advantages and risks arising from the agreement to process his personal (medical) data. In order to make this counterbalance, the data subject must be able to fully appreciate and understand the facts and implications of the data processing. In that awareness not only needs to be created about the consequences of giving consent, but also about the consequences of not giving consent. The information provided must furthermore be accurate and given in a transparent, clear and understandable manner. The use of vague or general phrasing is not considered to be in line with the law's requirements.

The duty to inform in medical regulations is conceived in a similar way, but encompasses more than providing the patient with enough information to consent in a well-considered way. The healthcare practitioners' duty to inform under medical regulation entails not only an obligation to provide the patient with a minimal level of information (as is the case in data protection regulation), but is really an active duty to make the patient understand the planned intervention. This implies that the health care provider will, for example, have to actively ask and answer the right questions instead of waiting passively until the patient asks them. In the end the patient should have complete knowledge about the proposed intervention. Therefore the healthcare professional has:

- to present the patient with a clear oversight of the intervention, including the painful or invasive nature thereof and its possible urgent character but without excessive medical technical details;
- to inform the patient on the duration of the treatment, number of necessary interventions and the frequency they will be effectuated with;
- to communicate all contra indications, possible side effects and risks of which he is aware that they could occur to normal patients in the same situation. The relevance of the risk should be judged according to the risk frequency, the gravity of the risk and the physical, mental and socio-economic disposition of the patient;
- depending on the balance of what the patient would like and would not like to know and what would and would not be useful for him in his specific situation "all other relevant clarifications" need to be provided, amongst which the total price of the treatment including refundable and non-refundable costs and the existence of a waiting list;
- Finally if several treatments are available, the practitioner is obliged to inform the patient about them and explain the pros and cons but without having to discuss every detail of it. It entails for example the difference in price, duration of the rehabilitation and health risks.

All these information elements are subject to one exemption: when the healthcare provider is of the opinion that informing the patient on his health situation would cause serious harm or disadvantage to the health of the patient, he can exceptionally forsake his obligation to inform. At least this is on the condition that the healthcare provider consults a colleague. It has however to be noted that this exception does not exempt the healthcare professional from informing the patient about the intervention itself. All the information necessary to understand an intervention has to be provided, only information regarding the health situation of the patient can be withheld.

The minimal information elements required to be provided under data protection regulation are as specifically regulated, but due to the different purposes of both types of consent, they are different. The patient or data subject needs provided with:

- the name and address of the person responsible for the data processing and - if there is one - his representative;
- the reason and purposes of the processing;
- the recipients or categories of recipients of the data;
- the list of authorized users or people who can access the data;
- the existence of the rights of the data subject, especially those to access and to rectify personal data;
- the legal basis on which the data are being processed.

Understandably the information for a medical consent needs – in principle – to be provided by the practitioner who proposes and will execute the treatment or operation. When several practitioners are involved, every practitioner should be giving the information related to his field of expertise. When an intervention is performed by a medical team, the obligation to inform lies however not with every team member, but with the practitioner who was chosen by the patient.

The information about the data processing on the other hand needs to be provided by the controller (the person responsible for the data processing). In certain cases the controller can also be the healthcare practitioner himself, but this will depend on the technical organization of the data processing.

2.1.4 Upfront

The information, whether necessary under medical regulations or under data protection regulation, needs to be provided at the latest before respectively the treatment is started or the data are being collected, but advisably in time for the patient to let it sink in and reflect on it. With regard to consent for treatment for example, the patient needs to be given enough time to take a well considered decision, to consult another practitioner for a second opinion or - in that particular case – to prepare his farewell from loved ones.

With regard to the consent for medical treatment it has furthermore to be noted that the patient must be informed at a moment he possesses of full awareness and adequate reflection capacities. This is not the case when s/he is drowsy due to drugs, when already on the operation table or when administered an anaesthetic.

2.1.5 Form

Whereas the consent for medical treatment and for data processing were more or less concurrent with regard to all of the previous requirements (free, specific, informed and upfront), their required form does differ. In the Belgian Data Protection Act it is explicitly stated that the processing of health information requires a written consent, but consent for medical treatment can under Belgian regulation be given orally.

Through the data protection's requirement of a written consent, the Belgian legislator opted for a stringent interpretation of the European Data Protection Directive, which only requires the consent to be explicit. This stipulation is often experienced as a hassle, but it should not be forgotten that also for the sake of legal evidence some kind of "written" act would be highly advisable to obtain. Now that solutions for electronic signatures become more widely available, it might moreover become easier to obtain a written consent fast, even though as we will discuss in the next section, this comes yet again with specific regulations.

The consent for medical treatment on the other hand has to be explicit but does not necessarily have to be in writing. If the healthcare provider can moreover unambiguously deduce the patient's consent from his conduct or what he says, an implicit consent can also be accepted. This can for instance be the case when the patient spontaneously exposes his upper-arm for the collection of blood. On request of the patient or the healthcare provider and with agreement of the other party, it is possible to record the consent in writing and add it to the patient's medical record. This request can for instance be made with view to legal evidence, anticipating possible discussion after the intervention. The information which has to be given to the patient under the duty to inform of the healthcare practitioner can be communicated orally too, but again the patient has the right to have it confirmed in writing. This is because a written copy allows the patient to read everything trough in a quiet setting. It does have to be noted that this written copy may however never replace the oral discussion.

2.2 Digital and digitalized patient consents

In the majority of countries specific legislation on the digital or digitalized patient consent is not available. Account therefore has to be taken of general evidence regulations, often including specific regulations on the electronic signature. In Belgium this is no different.

A distinction has to be made between a digital and digitized patient consent. When the consent has been given on paper, but scanned and stored in a digital way, it is called a digitized patient consent. If the consent is on the other hand immediately given in an electronic way e.g. through an online service, it is called a digital consent.

In the case of digitization the most pressing question is how to preserve the legal value of the paper consent as a piece of evidence. The answer to that question lies within the general principles of evidence law. The principle of providing "evidence" is all about the demonstration of the accuracy of a fact, or of the reality of a legal transaction in case of dispute. Belgian civil evidence law – applicable to all areas of law in so far as a contract or other law does not explicitly provide otherwise – is builds on a closed list of acceptable evidence. This means that only certain types of credible evidence are taken into account before court. The types of evidence accepted for the delivery of proof are: written evidence, the testimony of witnesses, circumstantial evidence, the parties' admission and the oath. These pieces are furthermore hierarchically organized. In that hierarchy the signed paper document is traditionally considered the by far most reliable because it a) allows identification of the author, b) has a guaranteed integrity and c) assures an appropriated content. When the paper document is digitalized by using for instance scanning techniques, these three functions (identification, integrity and appropriateness) are no longer fully guaranteed. Therefore the law of evidence traditionally considers a scan as a copy and not as an original document. Since a copy is only accepted as a piece of evidence when the original can still be produced, the digitized consent can only be submitted as a piece of evidence when the paper document is not destroyed. When digitalizing informed consent documents, the paper versions should thus be stored too in order to avoid every risk.

With regards to the digital informed consent, the question on how to sing was the most pressing. Since the introduction of electronic signatures however, an electronically signed document can be considered equivalent to a manual signature on paper. Two conditions determine the validity of an electronic signature. On the one hand it must be possible to attribute the electronic data that constitute the signature to a particular person. This reflects the document's functions of identification and appropriateness. On the other hand the electronic data constituting the signature must demonstrate the preservation of the document's integrity, reflecting the function of integrity verification. The European Directive on Electronic Signatures defines one type of electronic signature, the so called "qualified electronic signature", for which these conditions are assumed to be fulfilled. The "qualified electronic signature" is an "*advanced electronic signature, based on a qualified certificate and created by a secure signature creation device*". Electronic signatures based on other techniques can also be accepted before court, at least when they fulfil the same two conditions.

On the condition that it is accompanied by a legally valid electronic signature, the digital consent will thus automatically be accepted as a valid piece of evidence, also when it is given online. Consequently the digital informed consent will not only be time-saving but can also be considered a safer option in comparison with the digitalized consent. It does however have to be noted that the digital consent itself will also be qualified a processing operation and will need to meet the general data protection principles.

2.3 Consent to express approval and disapproval

Although at first sight – certainly in legal terms – consent is ought to be a positive reply to a request to process data, it could also express a negative attitude towards a certain processing action. A patient can, in other words, express his positive consent to allow certain processing actions, but also his negative consent to not allow certain processing actions. In this sense the negative consent should be regarded as a substitute for the right of the patient to express his wishes not to register certain information or not to share it with certain individuals or parties.

The use of negative consents is however not unlimited. In the legal doctrine it has been argued that a negative consent can not always be forced onto a healthcare professional, since the right of the patient not to share certain information with certain individuals, is not absolute. The criterion applicable when judging the patient's request not to register certain information or not to share certain information, must always be the impact thereof onto his treatment. The patient's negative consent can thus only be accepted when it does not force the patient's treatment into a tight corner. A balance will have to be found between the autonomy of the patient and the feasibility for the healthcare professional to care for the patient in the best possible way.

3. PRACTICAL IMPLICATIONS: A NEED FOR ADVANCED LIFECYCLE-MANAGEMENT

When collecting informed consents on a – in quantity and duration – large scale, it is important to realize that many different actors or parties become involved. Therefore it needs good management, life-cycle management. Within the Flemish Share4Health[2] project it was found that advanced life cycle management solutions can cope with the very specific, detailed and highly complex

[2] Share4Health was an interdisciplinary project funded within the Flemish IBBT network focusing on information sharing between different healthcare providers. IBBT (Interdisciplinary institute for BroadBand Technology) is a research institute founded by the Flemish Government in 2004 (more info: www.ibbt.be). The Share4Health project is a project in cooperation with the following companies and organizations: Agfa, Cronos, Custodix, KLAV, Corilus, SoSoeme, HealthOne and APB.

requirements set by legislation. It was furthermore found that through such an advanced life-cycle management solution a shift from the healthcare practitioner to the patient as driving force behind the consent process could be stimulated and patient empowerment could be enforced. In the following paragraphs the paper introduces the developed solution from a legal point of view.

Questions to be asked relate first of all to the actors involved. Who is responsible when managing informed consents on such a large scale? Who owns the consents? Secondly questions are asked with regard to the management of the consent: what are the requirements on the storage? Who can access the consent? Are they part of the electronic health record and do they need to be handled as sensitive data? A third question that arises is what to do when changes to the informed consent become essential? What if additional informed consents are required?

In answering these questions it is crucial to keep the purposes of the informed consent in mind: the consent as legal ground – and thus in fact as legal requirement – for either medical intervention or data processing and as crucial element of evidence in case of conflicts. Both these purposes cause the informed consent to be a rather static document which cannot easily be altered, modified or changed. As described above, a consent needs to be given a priori, for a determined purpose, for a specific timeframe and by a specific person to a specific person or group of persons. Following evidence law, a consent can furthermore only be considered valid when the document is signed. As also described in the legal discussion: if a consent is subject to change, evidence law will require a new signature in order for the document to preserve its validity.

3.1 Ownership over consent

A question posed very often is the question to the ownership of the consent. At least in legal terms, the answer is however short: no one. The patient or data subject keeps, in principle, control over his informed consents as they indicate his wishes. Legally ownership is however defined as: "*the right to enjoy an object to the fullest and dispose thereof […]*". Ownership is, in other words, a property right that entails you to dispose of an object in an absolute way. With regards to medical data or medical patient files, two problems arise. First of all medical data or medical files cannot be considered an object in the sense of Civil Law. Secondly and most importantly, no one nor the healthcare practitioner, patient, hospital or archiving party, can claim the absolute disposition of medical data or a medical file. A healthcare practitioner for example is not allowed to do whatever he wants with the collected data. On the contrary, he is bound by his professional secrecy and is not allowed to share the data without any justified reason. The patient on the other hand is not allowed to physically hold on to his (original) data or files. The data or files remain stored on the healthcare practitioner's computer, the computer of the pharmacist or in the computer system of the hospital, a company or a government[3].

Therefore the debate on ownership over consent is rather a pseudo-debate for the questions concerning the obligations to store and the rights to access consents.

3.2 Consent management

There are very little firm rules on the management of consent. We therefore take the legal qualification of the consent, its legal value in case of a need for evidence and the rights of the involved actors as guiding factors.

A first pressing practical issue is the storage of the consents. Who has to store them? For how long do they need to be stored? Are they considered part of the electronic health record and do they need to be handled as sensitive data? Many questions, but very little firm rules. In practice it will most often be the medical practitioner who stores the informed consent. It is after all him who needs to obtain the document from the patient and might need it as legal evidence when a discussion would arise. For the same reason it is advisable to store the consent at

[3] For the purpose of completeness it has to be added that academic, high level discussions on the desirability of property rights and intellectual rights over personal data are ongoing. It is said that the creation of a new property rights model will allow data subjects to sell their data and that this would enforce companies to think more carefully about what data they really need to collect and would force them to take better investment decisions. However, these theories are still very immature and therefore not further included in this paper.

least for the duration of what was consented for and to be on the safe side for 30 years, in concurrence the (Belgian[4]) legal requirements for health records. After this 30-year period, the healthcare practitioner is obliged to destroy the informed consent document, unless, the purpose of the consent would require it to be kept longer.

Although it is thus advisable to organize the storage of the patient consent concurrent to the storage of the patient health record, the question whether the patient consent should be part of the health record is highly debatable. For two reasons we argue that this is not the case. A first argument can be found in the concept of 'health record' as used by different legal texts and as defined by the article 29 working party. The art 29 WP defined an 'electronic health record' as: "*a comprehensive medical record or similar documentation of the past and present physical and mental state of health of an individual in electronic form and providing for ready availability of these data for medical treatment and other closely related purposes*". The expression 'closely related purposes' in that definition refers to article 8, 3) of the Data Protection Directive, which states that sensitive personal data can be processed "*where the processing of the data is required for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment or the management of healthcare services*". This expression does not refer to the informed consent, rather on the contrary. In Belgium the concept of 'health record' is not defined by any law, but along the same line, they too use the concept of health record always in relation with medical documentation. A second argument for not considering the informed consent as part of the health record is that the informed consent document does not serve the same purposes as a health record. The purpose of health records is double: it has to guarantee the continuity of care and it has to allow post factum control in medical disputes. Although the informed consent, both for medical treatment and data processing, also serves as a piece of legal evidence, it's purpose is not to proof that the right medical treatment was chosen, but that the patient's will was respected. Furthermore the purpose of the informed consent is not to guarantee continuity of care, but it is a prerequisite to the creation and use of an electronic health record and it has to inform the patient. It is consequently concluded that the informed consent document should be regarded as an administrative rather than a medical document and should not be considered as part of the health record.

The qualification of the informed consent document as an administrative rather than a medical document and therefore not part of the health record, does not exclude its qualification as sensitive (health) data. As discussed above, all data that have a "*strong and clear link*" with a person's health status are considered health data. Whether or not in an informed consent document this link is present, will have to be judged case by case. This is because the evaluation of that criterion will depend on the exact content of the informed consent document. If the document refers to a specific kind of treatment or it refers to a specific disease, it is quite likely that the informed consent document too will be qualified as sensitive. If, on the contrary, only very general terms are used, such as "*the proposed treatment*" it cannot be qualified sensitive.

3.3 Alternation and retraction of the informed consent

The informed consent documents will not only have to be stored, they will also have to be managed further. Alternations may be necessary and access may have to be provided. Amongst technical reasons, this is mainly because involved parties can enforce certain rights.

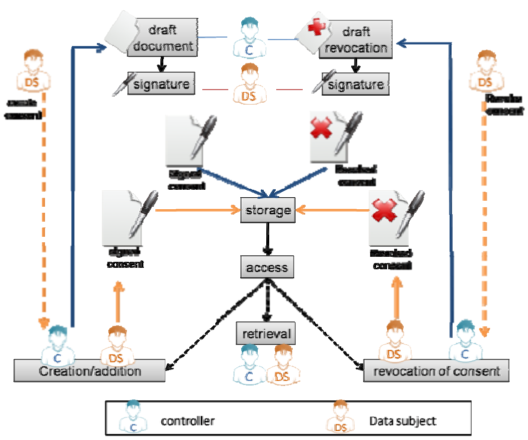
The data subject has first of all the rights to access the informed consent document and to retract his consent at all times. The retraction poses the most problems. Although the data subject can never demand the controller to assign a retroactive effect to his retraction, a retraction does entail that there is no longer a valid legal basis and no more data can be processed. Furthermore it entails that all the previously processed sensitive (health) data have to be deleted. Whether or not the informed consent document itself will also have to be deleted, depends – following the discussion above – on its qualification as sensitive or not. Because this consequence might however be pernicious for the controller's ability to proof the validity of earlier processing actions, it is not realistic to delete the full informed consent document. A compromise could be

[4] This term differs greatly amongst European Member States.

to only delete the sensitive information in the document or to replace the original informed consent by the agreement to retract.

The controller, secondly, drafts the informed consent, receives the informed consent after signature by the data subject en stores it for later use. Although, as described above, he cannot be qualified as the owner, he can be considered the keeper of the material document. The controller therefore has to be able to access the original document, delete it and alter it, but of course only when acting upon the patient's wishes. Consequently this implies that a controller who wishes to delete an informed consent, will have to verify if that has consequences for the data subject. If it concerns informed consent documents which were only kept for evidence purposes but do not have any value anymore for the data subject, it is considered fair that the controller can delete it without consulting the data subject. If however, the controller wishes to delete informed consent documents because he does not wish to continue the application for a specific subject, acting without consulting the patient can no longer be considered fair. The same goes for alternations of and additions to an informed consent document. In practice this means that a new signature will have to be obtained from the data subject in order for the altered informed consent to remain valid.

The figure below shows the different interactions expected in an advanced lifecycle management system.



© Kim Wuyts, 2010

The only problem the advanced life-cycle management solution could not resolve is the risk of a consent overload and/or lack of interest. Research on how to personalize informed consents and make them more appealing to patients by indicating individual risks for instance, is being carried out in medical centres in the US. Yet, as Michael Ende would say, this constitutes another research story.

4. CONCLUSION

The use of patients' consent is growing. In medical settings the patient's consent needs to be obtained before providing him/her with medical treatment and before processing his/her (health) data. Because these two consents stem from different regulations, respectively national medical regulations and the European Data Protection Directive, and serve different purposes, respectively inform about the healthcare intervention and about the data processing process, they are subject to different requirements. The basic principles they are subject to are the same, but their interpretation and implementation differs, causing a complex web of norms for the healthcare practitioner.

A consent is only regarded valid when given freely, before the action starts, after being informed, for a specific purpose and in a specific form. This implies that the consent first of all needs to be a voluntary decision, taken by an individual in possession of all of his faculties and in the absence of coercion of any kind. Secondly it implies that the patient's consent must be obtained for every medical intervention and each time a defined set of data is being processing for a delineated purpose by a determined individual or category of persons. Thirdly the patient's consent can not be obtained without adequately informing him on what he is consenting for. In the case of consent for treatment, the healthcare practitioner has an active duty to adequately inform the patient on the planned intervention. When it concerns a consent for data processing, a passive duty to inform the patient on the data processing process is imposed on the controller. The forth implication concerns the timing of the consent request. This needs to be upfront in any case, but advisably in time for the patient to let it sink in, reflect on it and consult relatives or other professionals. Lastly the consent must be obtained in the form as required by law: orally when it concerns medical treatment and in writing when it concerns data processing. It is however at all times the right of both parties to request that the consent for medical treatment is recorded in writing.

The use of consents in health environments is not only complex because of the detailed requirements imposed on them, also the rising amount of occasions on which a consent needs to be obtained is experienced as a hassle. When people – both patients and practitioners – are confronted with a consent-overload, the risk to a lack of interest and consequently a less empowering effect rises. The use of digitalized consents can be a first instrument to counter this. When moving towards digitalized patient consents account has to be taken of general evidence regulations and regulations on electronic signatures. The most pressing questions consequently concern how to preserve the legal value of the patient's consent and how to sign them? The principle of providing "evidence" is all about the demonstration of the accuracy of a fact, or of the reality of a legal transaction in case of dispute. Therefore it is better to use electronic means when obtaining the consent originally instead of making a digital copy of the original paper consent. Now that electronic signatures are furthermore accepted as equivalent to a manual signature, the last legal burden for digital patient consents has been removed.

A second instrument to overcome a loss of empowering effect of the patient consent is the use of an advanced lifecycle management application. Through such a solution both positive and negative consents could be expressed, the large quantity of consents can be better managed and becomes surveyable for both the patient and the healthcare practitioner and a shift from the healthcare practitioner to the patient as driving force behind process could be stimulated. The main problem to overcome has however to do with the static character of a consent. It cannot easily be altered, modified or changed. As described above, a consent needs to be given a priori, for a determined purpose, during a specific timeframe and by a specific person to a specific person or group of persons. Following evidence law, a consent can furthermore only be considered valid when the document is signed. As also described in the legal discussion: if a consent is subject to change, evidence law will require a new signature in order for the document to preserve its validity. It is therefore crucial to include all actors in the management solution and not only allow storage, but also retraction and addition. It is advisable to organize the storage of the consent in concurrence with the storage of electronic health records, even though the consent is not part of that record, with respect for all parties' rights. The data subject has the right to access the informed consent document and to retract it at all times. The controller drafts the informed consent, receives the informed consent after signature by the data subject en stores it for later use. The controller therefore has to be able to access the original document and delete it at least as long as he's acting upon the patient's wishes.

REFERENCES

Article 29 Working Party, 2007. Working Document on the processing of personal data relating to health in electronic health records (EHR).

Beaulieu, D., 2010. Updating informed consent in the era of patient-centered care. *FierceHealthcare*.

Cornelis, L., Simont, L., 1987. Bewijsrecht en Technische Evolutie: enkele overwegingen. *in* De Vroede, P., *Technologie en Recht.*

De Bot, D., 2001.Verwerking van persoonsgegevens. *Kluwer*.

Dijckhofz, W., 2003-04. Het recht op informatie en geïnformeerde toestemming. *T.Gez./Rev.dr.Santé.*

Dumortier J., Dekeyser, H., 2005. Digital Archiving: the new challenge?. *IRIS*.

Fagnart, J., 2006, Information du patient et responsabilité du médecin. Bruylant.

Graux, H., Dumortier, J., 2009, Privacywetgeving in de praktijk. *UGA Antwerpen*.

Leleu, Y., Genicot, G., 2001. Le droit médical : aspects juridiques de la relation médecin-patient. *De Boeck & Larcier*.

Mougenot, D., 2002. Droit des obligations La prevue. *Larcier*.

National Council of the Order of Physicians. 2005. Advies van de Nationale Raad van de Orde van geneesheren van 16 oktober 2004. *Tijdschrift Nationale Raad*.

Nys, H., 2005. Gneeeskunde, Recht en Medisch handelen. *Algemene Praktische Rechtsverzameling*.

Tack, S., Balthazar, T., 2007. Patiëntenrechten – Informed consent in de zorgsector: recente evoluties, *C.A.B.G*.

Vansweevelt, T., Swennen, F., Ter heerdt, J., Weyts, B., Jeger, N., 1997. Het voorontwerp van de wet patiëntenrechten: een kritische analyse. *T.Gez./Rev.dr.Santé*.

Legislation:

R (97) 5 on the Protection of Medical data, European Council, 13 February 1997.

Directive 95/46 on the Protection of Individuals with regard to the Processing of Personal Data and the free movement of such data, 24 October 1995.

Directive1999/93 on a Community Framework for Electronic Signatures, 13 December 1999.

Belgian Act on Patient Rights of 22 August 2002 and Proposal of law concerning the rights of the patient, M.v.T., *Parl.St.* Kamer 2001-2002, 1642/001, 24.

Belgian Act the protection of privacy in relation with the processing of personal data, 8 December 1998 and the on Royal Decree implementing the Law of 8 December 1998 on the protection of privacy in relation with the processing of personal data, 13 February 2001.

Belgian law introducing the use of telecommunication means and of the electronic signature in proceedings in and out of court, 20 October 2000.

Belgian law of 9 July 2001 on the establishment of certain rules relating to the Legal framework for electronic signatures and certification services, 9 July 2001.

Case Law:

Ontario High Court of Justice 1983, *Ferguson vs Hamilton Civic Hospitals*, 122 BLR (3d) 214.

Antwerpen 13 september 2005, *T.Gez./Rev.dr.Santé* 2006-2007, 117.

Antwerpen 30 april 2001, *T.Gez./Rev.dr.Santé* 2002-2003, 171.

Gent 11 maart 1992, *T.Gez./Rev.dr.Santé* 1995-1996, 54.

Rb. Antwerpen 12 mei 2004, *T.Gez./Rev.dr.Santé* 2005-2006, 221.

Rb. Leuven 10 februari 1998, *T.B.B.R.* 1998, 163.

Rb. Antwerpen 13 mei 1993, *T.Gez./Rev.dr.Santé* 1996-1997, 190.

CHALLENGES FOR THE EUROPEAN NEXT GENERATION BROADBAND

Mr. Arturo Vergara, Mr. Antolín Moral, Ms. Ana Olmos, Dr. Jorge Pérez
Signal, Systems and Radiocommunications Department, Universidad Politécnica de Madrid (UPM), Av. Complutense nº30 C-431, 28040, Madrid, Spain
{avergara, amoral, aolmos, jperez}@gtic.ssr.upm.es

Keywords: Broadband platforms and services, ICT public policy and regulation, next generation access networks (NGA).

Abstract: This article presents an overview of European NGA roll-outs compared with leader regions (i.e.: Asia-Pacific and US), shows the main risks faced by operators of different natures (i.e.: incumbents, alternatives, cable, etc.) and outlines the main commercial, regulatory and public policy challenges for the European next generation broadband.

1 INTRODUCTION

The rollout of new generation broadband, commonly known as Next Generation Access (NGA) networks has become one of the main political and strategic priorities of most developed regions of the world in recent years. The impact of the economic crisis has highlighted the need to promote improvements in the competitive and innovation capacity of the economy. The economy of the future will be knowledge-based and focused on the Internet, and based on networks that enable widespread superfast broadband access at competitive prices. As a result, a large number of regions have adopted measures and plans to promote NGA rollout in order to position themselves advantageously in anticipation of an end to the economic crisis, foster sustainable economic growth and create employment opportunities.

One example is the European Union's Digital Agenda initiative, which has set a goal of universal 30 Mbps broadband by 2020, with at least 50% of households subscribing to connection services of over 100 Mbps.

Similarly, the National Broadband Plan in the United States has set the goal of providing 100 million households with affordable broadband service access of at least 100 Mbps by 2020, with an interim objective of 100 million households with affordable 50 Mbps access services by 2015.

Other countries in the Asia-Pacific region have adopted direct infrastructure investment measures instead of promoting private sector rollout. A recent example of this is Australia, which is investing approximately €23 billion to cover 90% of the country's households with a Fibre-To-The-Home network.

The political pressure to achieve "100 Mbps" NGA network rollout is therefore clearly evident. However, the rollout of these networks is a complex process that not only depends on political decisions, but also market, regulatory and technical conditions. The path adopted by the various players towards next generation networks will not be homogeneous, nor will they respect technologies or platforms. Therefore, the commercial models and rollout strategies will follow the "one-size-does-not-fit-all" principle (Amendola and Pupillo 2008).

If the uncertainties associated with new network rollout outweigh incentives, it is likely that operators will show greater caution before undertaking substantial investment. This is the case of Europe, where a new infrastructure environment and framework needs to be built (EC, 2009a), and private operators have not yet acquired the strategies and sustainable business cases to support the upgrade of current broadband platforms (Noam 2010). European operators face a set of risks that erode their incentives to roll out NGA platforms. Although some European countries are planning to accelerate NGA rollout (Bourreau *et al.*, 2010), Europe is lagging behind Asia and the U.S. - the worldwide leaders in NGA penetration (see Figure 1).

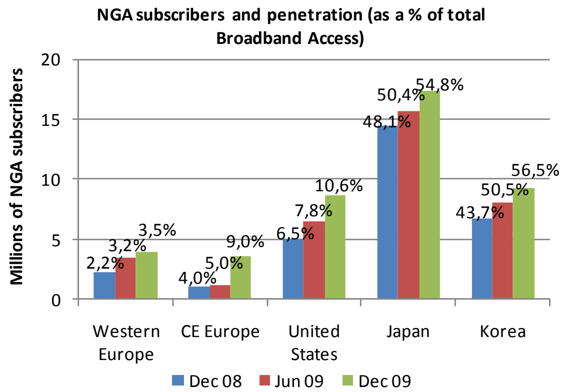


Figure 1: Global NGA subscribers and penetration.

If Europe does not want to fall further behind other regions in terms of ICT readiness, policymakers and industry players will need to work together to develop and align themselves with regulatory, public policy and supply models that foster investment in wide-scale NGA rollout, thus overcoming the challenges of next generation broadband.

The purpose of this article is to analyse the main obstacles to the development of fixed NGA in Europe and outline the challenges faced by the various players. The definition of NGA networks is that used by the EC (EC, 2009b), which includes fixed partial or complete fibre optic networks that enable greater capacity than existing "first generation" copper broadband networks. Generally, the definition includes rollout of Fiber-To-The-Home (FTTH) or Fiber-To-The-Building (FTTB); the evolution of copper networks to Fiber-To-The-Node (FTTN); and the upgrade of Hybrid Fiber Coaxial (HFC) networks through the deployment of DOCSIS 3.0 technology. The statistics used in the article have been obtained from IDATE (IDATE, 2010), as well as corporate information that has been made publicly available by the various operators involved.

Section 2 provides an analysis of the current state of NGA rollout in Europe. Section 3 discusses the key commercial challenges, while Section 4 addresses the regulatory challenges and Section 5 the role of government authorities. Finally, Section 6 presents the main conclusions.

2 EUROPEAN NGA SITUATION

By the end of 2009, the global market for NGA networks had reached a total of 46.5 million users globally, with a heterogeneous structure in terms of regions and countries (see Figure 1). Asia is in the leading position as a result of the implementation, since 2000, of specific public policies to encourage the rollout of ultra-fast networks (ITIF, 2008). In second place is the United States, which has benefited from a deregulation process that started in 2004 and strong competitive pressure from cable operators.

Western Europe is ranked in third place in the world, with a total of 3.92 million subscribers in 2009, and coverage of approximately 50 million households. However, the differences in the compitive dynamics of the various markets, the costs of different platforms, different regulatory models, public policy and geographical conditions show that NGA network development in Europe has been uneven, in terms of the players and platforms present, and the scope and pace of rollout (see Figure 2).

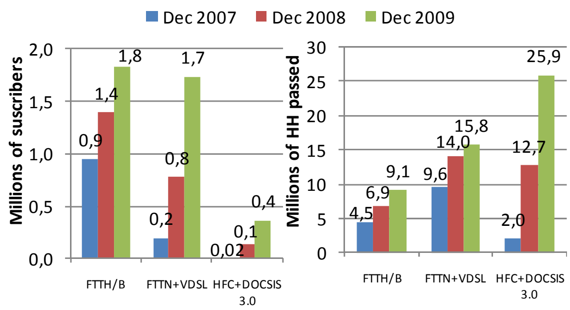


Figure 2: Evolution of NGA subscribers and HH passed

Each type of platform faces different issues depending on its cost structure, the availability of passive assets and the synergies that result from the move towards an NGA model.

Cable operators tend to be the first movers due to advantages on marginal cost and roll-out pace in adapting their networks to NGA in those areas where they already have coverage, as it is not necessary to modify their outside network plant.

For their part, established operators will tend to position themselves in a reactive manner, because of the high cost of developing NGA architecture and because they face regulatory asymmetry. This development, however, is a necessary step, as a result of the need to reduce the offer speed's dependence on the distance to the node, as well as the improvement of operational efficiency, and the necessary modernisation of networks.

Meanwhile, alternative operators based on local loop unbundling will tend to use wholesale products. However, in areas in which they reach a sufficient market share, they will be able to rollout their own access networks (e.g. the cases of Iliad-Free and SFR-Neuf in Paris).

In terms of rollout already carried out, the 2008-2009 period saw the completion of FTTN network rollout by operators such as Swisscom, DT and Belgacom, providing coverage for almost 16 million households. At the same time, there have been some delays in developing FTTH network rollout plans in the Netherlands and France due to regulatory uncertainty and the suspension of Dansk Fibernet's rollout in Denmark as a result of the purchase of its network by the established operator TDC.

However, there has also been a significant growth in FTTH networks in Germany instigated by regional utility companies (M-Net, NetCologne or Wilhelm Tel), as well as the beginning of FTTN rollout by Telecom Italia or FTTH by British Telecom. For their part, cable operators have upgraded a large part of their network coverage with DOCSIS 3.0 technology, which provides services with speeds of over 50 Mbps.

As shown in Figure 3, to date, NGA coverage levels of over 40% have only been achieved in countries which, in the last decade, decided to completely adapt their networks to FTTC models (Belgium and Switzerland) or those with significant DOCSIS 3.0 rollout (Netherlands, Portugal, Switzerland and the UK).

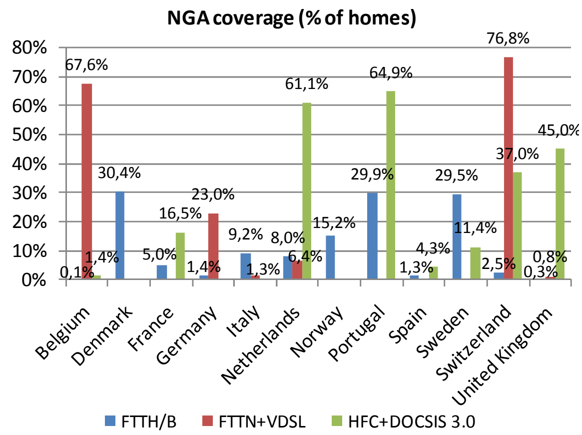


Figure 3: NGA coverage in Europe for different broadband platforms

In terms of the number of subscribers, growth in 2008-2009 stood at 70%, with a sharp fall in the second half. The main impetus for growth was a result of increased subscriptions in FTTN networks in Germany and Switzerland and sustained growth in FTTH networks in Denmark, Norway, the Netherlands, Sweden, France and Portugal. For its part, DOCSIS 3.0 network rollout by cable operators began to gain market share in France, Portugal and the United Kingdom. As Figure 4 shows, penetration levels above 10% were only achieved in Nordic countries (Sweden and Norway) and Switzerland.

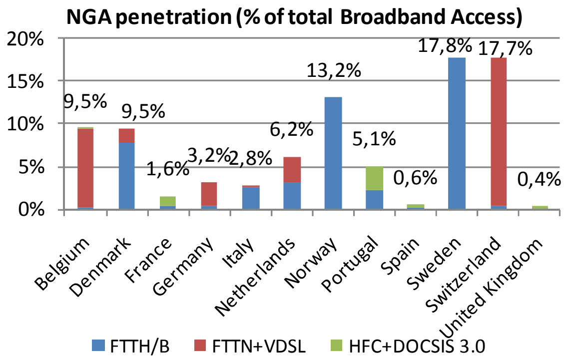


Figure 4: NGA penetration in Europe for different broadband platforms

The existence of NGA rollout stimulation dynamics (Vergara, 2008), such as: (i) competitive pressure from cable (Netherlands, Portugal and Switzerland); (ii) regional public utilities (Germany, Denmark, Norway and the Netherlands); or (iii) public policy (Sweden) have facilitated a real boost in rollout. In countries where these dynamics do not exist, late and limited rollout has occurred, with significant delays in rollout plans (France, Italy, the UK or Spain). In such cases, the existence of significant rollout investment planning is linked to: (i) the establishment of a regulatory model that promotes investment, (ii) the adoption of public policies that help and foster NGA rollouts in current non-profitable areas, the setting up of sustainable business models by operators, and the presence of alternative platforms that improve the competitive landscape (Vergara, 2009).

3 COMMERCIAL CHALLENGES

One of the main challenges for NGA rollout is the evolution of operator business models, traditionally based on access and telephony, which are facing structural stagnation as a result of the transference of the value outside the network to devices and services. Investment in NGA network rollout could improve this trend if operators manage to monetise the capabilities of new networks.

NGA has a great capability to provide new services, especially in video and storage. However, there are doubts about the ability of operators to significantly increase their income in order to justify NGA network rollout (Yankee Group, 2009). In the absence of disruptive ultra-wide broadband services, in the next 3-5 years, operators will gradually increase the speeds offered to the household market and increase the penetration of video-based services (Liberty Global, 2009).

Secondly, current business models based on flat rates, in which all users, regardless of their usage profile, enjoy the same pricing schedule, cast doubts on the ability of operators to create new business models and additional income from new networks. The traffic growth trend (CISCO, 2010), encouraged by this pricing and the emergence of intensive applications in bandwidth (such as Internet video and P2P networks), could threaten network capacity and current connection models.

In an environment of growing traffic demand, such as that envisaged by NGA rollout, operators must find new business models that are attractive to users and go beyond flat rates and the best-effort model. These models must be based on the ability of operators to offer different services with varying degrees of quality to users with different requirements. Capability, quality and security requirements are not always compatible with flat rate and best-effort models, which should reflect a differentiated service offering.

Thirdly, if the operators who invest in new networks can expect income only from connectivity, and most of the revenue-gain value created by NGA is capitalised by other agents, the feasibility of investment in new networks could be called into question. This leads to the net neutrality debate, which concerns the positioning of the various players in the value chain and the distribution of value in the Internet platform. The ability of operators to offer differentiated services through agreements with other application and content providers is perceived as a new source of revenue associated with new products (Vodafone, 2010). These benefits improve NGA rollout business cases and enable a more sustainable model to be achieved, in which not all costs and investments are exclusively assumed by operators (Faulhaber, 2009), and can be shared among agents. In order for such a paradigm to be implemented, operators need to be given the freedom to negotiate, an issue that is currently under discussion in Europe, where the Commissioner Neelie Kroes (EC, 2010) has formulated a preliminary set of principles in support of innovation, investment in new networks and the maintaining of the openness of the Internet.

Finally, it is necessary to consider the impact of the development of mobile technologies, which, in spite of being more limited in speed and capability, can be provided at lower end prices, and have the added value of mobility. It is expected that there will be further segmentation of users of new networks and services that give rise to different profiles. Major consumers of services and capability will be connected via fixed networks, while others with less requirements and more interest in mobility and nomadism will be focused on mobile networks.

4 REGULATORY CHALLENGES

In the face of current uncertainty about the viability of NGA business models, the lack of a clear, stable and predictable regulatory model, which would provide legal certainty for investors, is holding back the rollout of new networks. Although sector-specific regulation does not directly affect the return on investment, its indirect influence in terms of establishing obligations and access pricing largely defines the ability to establish such business models, especially for the established operators who find themselves in an asymmetric situation.

Unlike the U.S., where there is a high degree of competition between platforms (mainly through cable), in Europe, the competitive dynamic of the access market is based on intra-platform competition, which hinders the establishment of a model that maintains the balance between the degree of investment and competition (Siciliani, 2010). The debate on NGA regulation has been ongoing in Europe since 2008 with the first Commission consultation, which was revised in 2009 (EC 2009a) and is pending a final recommendation in the second half of 2010.

The EC proposes the application of consistent measures in the context of Market 4 (wholesale access to infrastructures) and 5 (wholesale broadband access), in order to ensure that competitors have *ex ante* wholesale access, with cost orientation, to infrastructures (access to civil works, sub-loop, etc.) belonging to operators with significant market power (SMP), while retaining access to services (active access) as a safeguard for competition. This model involves extending the current regulatory approach, based on the investment ladder, to new networks (Cave, 2010). As well, the EC's long-term obligations include measures for complete loop unbundling, for which it proposes a multi-fibre rollout model. As a mechanism for incentivising investment, the EC has proposed to include a risk premium and a degree of flexibility for duration and contracted volume (referred to as diversification or risk sharing) in the pricing. The model seeks to promote multi-fibre rollout and joint investment by eliminating cost orientation in those cases that permit access by other operators on an equal footing.

However, these measures have been criticised by regulators (ERG, 2009) and operators (ETNO, 2009) for being too prescriptive and detailed. The general imposition of cost orientation in all wholesale services, despite the risk considerations, places a disproportionate burden on some infrastructures that do not yet exist, and can reduce the incentives for rollout and limit the establishment of a sustainable competitive model across alternative platforms.

The regulation should include a realistic scenario that does not impose restrictions inherited from existing networks and is in line with the new competitive market dynamic. The flexibility in the technologies used, the geographical diversity and scope of the rollout should be basic principles in regulation aimed at incentivising the rollout of new networks.

5 POLICY CHALLENGES

Government could play a key role in developing specific public policy to foster the rollout of next generation fixed and mobile access networks. These initiatives need to tackle the inhibiting factors and obstacles with different mechanisms that can range from providing civil works in public spaces, applying financial incentives for the rollout of new networks, facilitating funding and public/private partnership, to direct investment in infrastructure. Governments can also become major consumers and providers of advanced services, and can undertake demand policies by expanding their own electronic services.

Broadband has a major impact on the economy and the development of a country. Although it is still not possible to state that the new generation of broadband networks will have a similar impact, governments have to confront the challenges posed by new networks so as not to be left behind. Faced with a situation of delayed rollouts, several European countries have proposed the financing of parts of the infrastructure, on a national level (e.g. Greece, Portugal), and on a local level (e.g. France, Netherlands).

In anticipation of these initiatives, in 2009, the EC published directives for state aid in the rollout of broadband networks. These directives divide the territory into three types of regions according to planned NGA rollout over the next three years. In this way, depending on the characteristics of geographical areas (population density, dispersion, income, housing type, etc.) and the market itself, a country can be divided into three types of area (EC, 2009b): (i) low profitability high risk regions, which cannot support sustainable NGA rollout by private industry (white areas); (ii) high profitability regions that can support the rollout of several NGA networks (black areas); and (iii) intermediate regions, in which the rollout of a single NGA infrastructure is usually the only feasible option (grey areas). The directives authorise public investment in infrastructure rollout in white and grey zones (under a larger set of constraints).

Although the EC directives provide a framework that respects private investment, it is possible that private investment will not be expected either in profitable areas, since three years is a very short time to plan and take the investment decision given current uncertainties. If such public action is not carried out in coordination with private operators, the result could be the creation of small networks that have no realistic chance of viability in the medium term, which would lead to a consolidation process similar to what occurred with telephone and cable networks, becoming an additional burden on the sector.

Governments face the challenge of establishing balanced, flexible and intelligent stimulation plans, for which, consultation and agreement with the different agents, especially with operators and private infrastructure investors, is required.

6 CONCLUSIONS

Next-generation network rollout faces different challenges that regulators, government, policy makers, operators and civil society must address to stimulate and develop NGAs. The main challenges identified in this article are summarised below:

- Effective NGA rollout requires the establishment of a virtuous cycle to improve the investment climate through sustainability in the creation of value by the various players in the sector.
- Operators must work towards the establishment of new business models and pricing schemes that will enable them to monetise the value of new networks.
- The distribution of value generated by the Internet platform is a key element in NGA development. Relationships between players should be based on their ability to reach agreement, innovate and establish new business models with respect to an open Internet model. To do so, Europe needs to resolve the debate on Internet neutrality.
- The EC should establish a clear, stable and predictable NGA regulatory model, by providing legal certainty to investors. The Recommendation on regulated NGA access, planned for the second half of 2010, represents a key opportunity.

- The developed model should promote infrastructure competition, free up bottlenecks and enhance value generation. It should also facilitate investment schemes and shared risk for operators in areas where it is deemed appropriate.
- Public policy to promote NGA network rollout should be strengthened in the form of national plans through agreement with operators.
- Public incentives should be focused on maximising public funding in grey and white areas, as well as fostering competition in the long run amongst operators with their own infrastructures.

ACKNOWLEDGMENTS

El primer autor quiere agradecer a la Universidad Politécnica de Madrid su financiación a través de su programa propio FPI.

REFERENCES

Amendola, G. B., Pupillo, L. 2008. The economics of next generation access networks and regulatory governance: Towards geographic patterns of regulation. *Communications and* Strategies, vol. 69, pp. 85 – 105.

Bourreau, M., Cambini, C., Hoering, S., 2010. National FTTH plans in France, Italy and Portugal. *Robert Schuman centre for advanced studies.* EUI Working Papers, RSCAS 2010/51.

Cave, M., 2010. Snakes and ladders: Unbundling in a next generation world. *Telecommunications Policy, v*ol 34, issues 1-2, pp. 80 -85, February-March 2010.

CISCO, 2010. *Cisco Visual Networking Index: Forecast and Methodology, 2009–2014*. June 2010.

EC, 2009a. European Commission, *2nd Draft Commission Recommendation on regulated access to Next Generation Access Networks (NGA)*. Brussels.

EC, 2009b. European Commission, *Community Guidelines for the application of State aid rules in relation to rapid deployment of broadband networks*. Brussels.

EC, 2010. *Net neutrality in Europe,* Neelie Kroes speech address at the ARCEP Conference (L'Autorité de Régulation des Communications Electroniques et des Postes), Paris, April 13, 2010.

ERG, 2009. *I/ERG Response to the draft NGA Recommendation*. ERG (09) 16rev3. July 2009.

ETNO, 2009. *ETNO Reflection Document in response to the Commission Recommendation on regulated access to Next Generation Access Networks (NGA) – RD307*, Brussels.

Faulhaber, G., Farber, D. 2009, *Innovation in the Wireless Ecosystem: A Customer-Centric Framework, FCC Dockets GN 09-157 and GN 09-51*, Sept 30.

IDATE, 2010. *FTTx Watch Service 2010 - Database Markets & Forecast*, Montpellier, May 2010.

ITIF, 2008. *Explaining International Broadband Leadership.* The Information Technology and Innovation Foundation (ITIF). Washington, May 2008.

Liberty Global, 2009. Next Generation Competition - Driving Innovation in Telecommunications. *Liberty Global Policy Series*. October 2009.

Noam, E., 2010. Regulation 3.0 for Telecom 3.0. *Telecommunications Policy, v*ol 34, issues 1-2, pp. 4 - 10, February-March 2010.

Siciliani, P, 2010. Access regulation on NGA–A financial, market-led solution to bridge the gap between US and European diverging regulatory approaches. *Telecommunications Policy, v*ol 34, issues 5-6, pp. 287 -298, June-July 2010.

Vergara, A., Pérez, J., 2009. Competencia entre plataformas alternativas de acceso. *Regulatory and Economic Policy in Telecommunications Review*. vol 3, pp. 94 – 106. Madrid, September 2009.

Vergara, A., Pérez, J., Moral, A. 2008. Comparative analysis of operators' strategies for the rollout of next generation access infrastructure in European markets. *Proc. of 17th biennial conference of the International Telecommunications Society (ITS)*, Canada.

Vodafone, 2010. The Economics of the Internet. *The Vodafone Policy Paper Series*. Number 11, April 2010.

Yankee Group, 2009. *Next Generation Access Services Analysis of Portfolios Yankee 2009. FTTH Council*. Copenhagen, September 2009.

SELF-MANAGEABILITY IN THE CONTEXT OF NEXT GENERATION NETWORKS

Ioannis P. Chochliouros[1], Anastasia S. Spiliopoulou[2], Maria Belesioti[3],
Evangelos Sfakianakis[4], George Agapiou[5], George Diakonikolaou[6],
Nancy Alonistioti[7], Panagiotis Spapis[8], Apostolos Kousaridas[9] and Andrej Mihailovic[10]
[1,2,3,4,5,6]Hellenic Telecommunications Organization S.A. (OTE), Research Programs Section,
99, Kifissias Avenue, 15124 Maroussi-Athens, Greece
ichochliouros@oteresearch.gr, aspiliopoul@ote.gr, mbelesioti@oteresearch.gr,
esfak@oteresearch.gr, gagapiou@oteresearch.gr, gdiako@oteresearch.gr
[7,8,9] National and Kapodistrian University of Athens,Dept. of Informatics and Telecommunications,
15784 Panepistimiopolis, Ilissia, Athens, Greece
nancy@di.uoa.gr, pspapis@di.uoa.gr, akousar@di.uoa.gr
[19]King's College London, Division of Engineering, Strand, WC2R 2LS, London, United Kingdom
andrej.mihailovic@kcl.ac.uk

Keywords: Autonomicity, cognitive networks, Future Internet (FI), networking awareness, Network Management (NM), monitoring protocols, Next Generation Networks (NGNs), self-adaptation, self-management, service awareness.

Abstract: The work examines several critical aspects from the potential inclusion of autonomicity and self-manageability techniques-methodologies in the scope of the deployment of next generation networks (NGNs) comprising an essential part of the Future Internet (FI). In particular, we first identify the significant strategic importance for further growth of modern NGN/FI infrastructures, to provide a variety of innovative facilities-services, decisive for further evolution of the market. Then we discuss some major challenges for the future networks, among which is the option for effective network management. FI should have a considerably enhanced network manageability capability, and be an inseparable part of the entire network entity. In the scope of evolving NGN architectures we further evaluate the vision for developing self-managing networks, able to constantly adapt their configurations and optimize their performance, by incorporating capacities of self-management/self-adaptation, as well as survivability in uncertain and dynamic environments. We further discuss a new network manageability paradigm, based on the original context of the Self-NET research project effort that engineers the FI based on cognitive behaviour with a high degree of autonomy and we incorporate some recent findings of the project. We also discuss benefits for the market actors and we conclude with some essential concerns.

1 INTRODUCTION

There is an extensive consensus that the Internet, *as one of the most critical infrastructures of the 21st century,* can critically affect traditional regulatory theories as well as existing governance practices (Commission of the European Communities, 2009a) But, as the future of the Internet comes into consideration, in parallel with the appearance and/or the development of next generation networks (NGNs), even greater challenges appear, with many concerns relevant to privacy, security and governance and with a diversity of issues related to Internet's effectiveness and inclusive character (Reding, 2009). With its several billion users worldwide, the present Internet is a world-wide success in terms of connecting people and supporting communities in general. It is a universal integrated communications network and service platform (also supporting the fabric of the European economy and European society). Future related facilities will "attract" more users to innovative services requiring greater mobility and bandwidth, higher speeds and improved interactivity through the launch of many interactive media- and content – based applications (Chochliouros and Spiliopoulou, 2005). Nevertheless, such claims necessitate a more secure, reliable, scalable and easily manageable Internet architecture. If effectively deployed, the *Internet of the future* can bring novelty, productivity gains, new markets and growth. The large-scale integration of various technologies (i.e. "distributed" computing platforms, peer-to-peer services, diverse broadband access networks, mobile devices and sensors), has led to a reconsideration of the relevant underlying architecture that was conceived more than thirty years ago.

In fact, innovative functionalities with more enhanced performance levels are necessary to sustain the real-time requirements of a multitude of novel applications. Furthermore, the Internet underpins the whole global economy. The networking effect has made possible an accelerated and universal diffusion of innovation. The diversity and sheer number of applications and business models supported by the Internet have also largely affected its nature and structure (Timmers, 1998). The Internet is evolving both in its use and in its technology: Born from the vision to create an "open infrastructure" to network computers across the world, Internet has become a socio-economic backbone of our society, with countless private and business users as well as governments relying on it on a daily basis. The "drivers" for this evolution are a mixture of emerging players with diverse and potentially changing interests, be it users, operators, manufacturers, service and content providers, together with advances in technology that have become available over the years (Future Internet Assembly-FIA, 2009).

1.1 The Strategic Scope for the Future Internet's Development

The ***Future Internet (FI)*** and all modern NGNs will not be "more of the same", but rather "appropriate entities" incorporating new technologies on a large scale that can unleash novel classes of applications and related business models (Afuah and Tucci, 2000). If today's Internet is a crucial element of our economy, FI will play an even more vital role in every conceivable business process. It will become the productivity tool "*par excellence*". At present, there are many so called "*Future Internet*" initiatives around the world working on defining and implementing a new architecture for the Internet intended to overcome existing limitations mostly in the area of networking (European Future Internet Portal, 2010). Beyond technological issues, the restructuring of business and social interaction processes unleashed by such (Internet-based) infrastructures could provide European stakeholders with a golden opportunity to lead a new wave of innovation and to establish a position in the Internet economy that is commensurate with their technological and scientific know-how (Blumenthal and Clark, 2001). The complexity of the Future Internet, bringing together large communities of stakeholders and expertise, requires a structured mechanism to avoid fragmentation of efforts and to identify goals of common interest. Appropriate action is therefore invaluable to pull together the different initiatives, in order to provide more potential options and/or opportunities for the market players involved.

Europe remains a international force in advanced information and communication technologies (ICT) and has massively adopted broadband and Internet services (Commission of the European Communities, 2009b). The European Union (EU) is actually a potential leader in the FI sector (Tselentis et *al.*, 2009). Leveraging FI technologies through their use in "smart infrastructures" offer the opportunity to boost European competitiveness in emerging technologies and systems, and will make it possible to measure, monitor and process huge volumes of information. This can also give the means to "overcome" fragmentation and to construct a related critical mass at European level, while fostering competition, openness and standardisation, involving consumer/citizen, ensuring trust, security and data protection with transparent and democratic governance and control of offered services as guiding principles. The current policy environment for the Internet-based economy is affected by three essential trends, i.e. convergence, creativity and confidence. With **convergence,** the Internet develops into a global platform to offer voice, video and data services, in a context where wireless and mobile are increasingly playing a larger role. Internet's "reach" is further expanding as it extends to include new devices and technologies. Secondly, the Internet has offered to our economies and societies great opportunities to exploit **creativity** and innovation with new applications, content and services for business and for consumers. Thirdly, today's increased dependence on the Internet for all manner of economic and social activity implicates that it is essential to enlarge the **confidence** of users and consumers (OECD, 2008). To be effective the

Internet requires an enabling environment for the related e-communications markets (Chochliouros and Spiliopoulou, 2003).

2 AUTONOMICITY AND SELF-MANAGEMENT FEATURES IN THE NGN DESIGN

2.1 Network Manageability Concerns

Innovation, the foundation for economic development, depends on rapid scientific advances. The face of the Internet is continually changing, as new services appear and become globally noteworthy, while market actors are adapting to these challenges through suitable business models (Commission of the European Communities, 2008). The current Internet has been founded on a basic architectural premise, that is: *a simple network service can be used as a "universal means" to interconnect intelligent end systems* (Galis et *al.*, 2008). Thus, it is centred on the network layer being capable of dynamically selecting a path from the originating source of a packet to its ultimate destination, with no guarantees of packet delivery or traffic characteristics. The continuation of simplicity in the network has pushed complexity into the end-points, thus allowing Internet to reach an impressive scale in terms of inter-connected devices. However, while the scale has not yet reached its limits, the growth of functionality and the growth of size have both slowed down. It is now a common belief that the current Internet is reaching both its architectural capability and its capacity limits (i.e. addressing, reachability, new demands on quality of service (QoS), service/application provisioning, etc.). The next generation network architecture will fix the shortcomings of the current Internet, including several significant aspects such as *security, privacy, trust and identity management*. It will have "hooks" for business and incentive models, support for semantics, support for mobility, and it will be resilient. This architecture will be flexible enough to support a range of application visions and business models in a dynamic way, ensuring convergence between technology, business and regulatory concerns. New communication technology will permit increasing connectivity, through both wired and wireless communication, both near-range and far-range. Enhanced communication services will open many possibilities for innovative applications that are not even envisioned today. Challenges for the Network of the Future may refer to a great variety of factors, including but not limited to: Dependability and security; scalability; services (i.e.: cost, service-driven configuration, simplified composition of services over heterogeneous networks, large scale and dynamic multi-service coexistence, exposable service offerings/catalogues); monitoring; Service Level Agreements (SLAs) and protocol support for bandwidth (dynamic resource allocation), latency and QoS; automation (e.g. automated negotiation/instantiation), and; the option for ***autonomicity***.

The resolution of these challenges would bring benefits to infrastructure-network and/or to service-application providers, in terms of: Simplified contracting of new business; establishing/identifying reference points for resource allocation and re-allocation; enabling flexibility in the provisioning and utilization of resources; offering the ability to scale horizontally, and; providing a natural complement to the virtualization of resources *-by setting up and tearing down composed services, based on negotiated SLAs-* thus simplifying accounting and revenue tracking. This can also involve benefits for service providers/consumers, in terms of: Ready identification and/or selection of offerings; the potential to automate the negotiation of SLA Key Performance Indicators (KPIs) and pricing; reduced cost and time-to-market for composed services; scalability of composed services, and; flexibility and independence from the underlying network details.

In addition, a current trend for networks is that they are becoming service-aware. Service awareness itself has many aspects, including the delivery of content and service logic, fulfilment of business and other service characteristics such as QoS and SLAs and the optimization of the network resources during the service delivery. Thus, the design of Networks and Services is moving forward to include higher levels of automation, autonomicity, including self-management. Conversely, services themselves are becoming network-aware. Networking-awareness means that services are executed and managed within network execution environments and that both services and network resources can be managed uniformly in an integrated way. It is commonly acknowledged that the Future Internet should have a considerably enhanced ***network manageability capability***, and be an inseparable part of the network itself. Manageability of the current network typically resides in client stations and servers, which interact with network elements (NEs) via protocols such as SNMP (Simple Network Management Protocol). The limitations of this approach are reduced scaling properties to large networks, and the need for extensive human supervision and intervention. A new network manageability paradigm is thus needed that allows NEs to be autonomously interrelated and controlled, that adapts dynamically to changing environments, and that learns the desired behaviour over time. The effective design of monitoring protocols so as to support detection mechanisms critical for the elaboration of ***self-organizing networks*** has to be based on a clear understanding of engineering "trade-offs" with respect to local vs. non-local and aggregated information, *for instance*. (Possible techniques for realizing such protocols include distributed tree algorithms, gossip algorithms and stochastic models). In fact, several issues identified in current network infrastructures impose the need for the introduction of an innovative architectural design. More specifically, existing web-based service front-ends are based on monolithic, inflexible, non-context- aware user interfaces (UIs). Furthermore, the diversity of services as well as the underlying hardware and software resources comprise management issues highly challenging, meaning that currently, a diversity in terms of hardware resources leads to a diversity of management tools (distinguished per vendor). In addition, security risks currently present in network environments request for immediate attention. This could be achieved by building trustworthy network environments (as well as communication, computing and storage infrastructures) to assure security levels and manage threats in interoperable frameworks for autonomous monitoring. Another important factor necessitating the need for dynamic FI environments is the reduction of "time-to-market", referring to the provision of services designated for the end users.

2.2 The Vision of a Modern Self-Managing Network

The future vision is that of a ***self-managing network*** whose nodes/devices are designed in such a way that all the so-called traditional network management functions, defined by the "*FCAPS*" management framework (Fault, Configuration, Accounting, Performance and Security) (ITU-T, 2000), as well as the fundamental network functions such as routing, forwarding, monitoring, discovery, fault-detection and fault-removal, are made to automatically "feed" each other with information (i.e. a sort of "knowledge") such as goals and events, in order to effect feedback processes among the different functions. These feedback processes allow reactions of various functions in the network (also including its individual nodes/devices), to achieve and maintain well-defined network goals (Pastor-Satorras and Vespignani, 2004).

Self-management capabilities may relate to a great variety of significant issues, such as: (i) Cross-domain management functions, for networks, services, content, together with the design of cooperative systems providing integrated management functionality of system lifecycle, self-functionality, SLA and QoS; (ii) Embedded management functionality in all FI systems (such as: in-infrastructure/in-network/in-service and in-content management); (iii) Mechanisms for dynamic deployment of new management functionality without interruption of actually running FI systems; (iv) Mechanisms for dynamic deployment of measuring and monitoring probes for services' and network's behaviour, including traffic. This also implicates SLA-aware sensing and continuous monitoring of systems' adaptations, together with the use of service monitoring in support of the self-management functionality; (v) Mechanisms for conflict and integrity-issues detection and resolution across multiple self-management functions; (vi) Mechanisms, tools and methodology construction for the verification and assurance of different self-capabilities that are "guiding systems" and their adaptations, correctly; these can also relate to mechanisms for allocation and negotiation of different available resources; (vii) Increased level of self-awareness, self-knowledge, self-assessment and self-management capabilities for all Future Internet systems/services/resources; (viii) Increased level of self-adaptation and self-composition of resources to achieve effective, autonomic and controllable behaviour; (ix) Increased level of self-contextualization and context-awareness for network and service systems and resources; (x) Increased level of resource management, including discovery, configuration, deployment, utilization, control and maintenance; (xi) Self-awareness capabilities to support system-level objectives of minimizing system life-cycle costs and energy footprints; (xii) Orchestration and integration of management functions, i.e. a service-driven dynamic orchestration, and; (xiii) Capabilities for the control relationships between self-management and self - governance of the Future Internet.

In such an evolving environment, *it is required the network itself to help detect, diagnose and repair failures, as well as to constantly adapt its configuration and optimize its performance*. Looking at ***Autonomicity and Self-Manageability,*** autonomicity (i.e. control-loops and feed-back mechanisms and processes, as well as the information/knowledge flow used to drive control-loops), becomes an enabler for self-manageability of networks (Dobson et *al*., 2006). Furthermore, new sensor technologies and wireless sensor networks provide options for inclusion of additional intelligence and the capability, for the network elements and/or domains to "*sense, reason and* *actuate*". Suitable systems with communication and computational capabilities can be integrated into the fabric of the Internet, providing an accurate reflection of the real world, delivering fine-grained information and enabling almost real-time interaction between the virtual world and real world. In particular, autonomous self-organizing systems are beginning to emerge and to be widely established (Boccalettia et *al.*, 2006). Such systems "*can adapt autonomously*" to changing requirements and reduce the reliance on centrally planned services, *especially if they are effectively joined with new network management techniques.* Operators may use these tools to guarantee QoS service in a period of exploding demand and rising network congestion at peak times. The trend in building dependable real-life systems and smart infrastructures today is "*to move from monolithic, centralized and strictly hierarchical systems to highly distributed networked systems with local and global autonomy*". When they are deployed in complex processes and services, these new systems exhibit promising features and capabilities such as modularity and scalability, low cost, robustness and adaptability. Some of the challenges for operators and service providers include management (especially in self-organized wireless environments), resilience and robustness, automated re-allocation of resources, abstractions of the operations in the underlying infrastructure, QoS guarantees for bundled services and the optimization of operational expenditures (OPEX). The requirement of a single, scalable and configurable architecture is an essential one of the driving forces for the FI. The variety and heterogeneity of the emerging business models, as well as the dynamic service composition and provision may lead to a situation of many Internets, with different architectural structure, requirements and functionality. Such a scenario will result in a nightmare of maintenance efforts, increased costs, incompatibilities and the like. It is thus important to try to build a single core architecture that maintains properties like configurability, extendibility, scalability and openness. Keeping the core architecture as generic as possible will offer the possibility to easily extend and adapt it to the requirements of the edge. Such a design will follow the rising trend of moving intelligence to the edge of the network.

Nowadays computing systems are open systems evolving in a dynamic complex environment. They are designed as sets of interacting components, highly distributed both conceptually and physically. The growing complexity of these systems and their large scale distribution make the use of traditional approaches based on hierarchical functional decomposition and centralised control no more applicable. Increasingly, a real need for new paradigms, mechanisms and techniques allowing endowing these systems with the capacity to autonomously manage their functioning and evolution, is expressed. Several among the existing technology systems are desired to sufficiently "exhibit" interesting characteristics, such as robustness, capacity of self-management and self-adaptation, as well as survivability in uncertain and dynamic environments.

Ubiquitous and self-organizing systems are not only disruptive technologies that impact the way how market actors organize core processes as well as existing structures in value chains and industry, but have also considerable impact (Kim et *al.*, 2008). The present Internet model is based on clear separation of concerns between protocol layers, with intelligence moved to the edges, and with the existent protocol pool targeting user and control plane operations with less emphasis on management tasks (Clark et *al.*, 2003). The area of FI is considered as a representative example of a "*complex adaptive organization*" (or "*entity*"), where the involved partners have conflicting goals and tension to maximize their gains. There is a need for new ways to organize, control and structure communication systems, according to new management schemes and networking techniques without neglecting the advantages of current Internet. Among the core drivers for the Future Internet are increased reliability, enhanced services, more flexibility, and simplified operation. The latter calls for including ***Network Management (NM)*** issues into the design process for FI principles. In general, NM is a service (or application) that employs a diversity of tools, applications, and devices to assist human network managers in monitoring and maintaining networks. Thus, network management should be an integral part of the future network infrastructure. Management is a key factor in manageability, usability, performance, etc., and is an important factor to the operational costs of any "network entity". FI requires a new management approach, promoted mainly by the necessity of support interoperability between heterogeneous, complex and distributed systems. In addition, FI should remain open for further and continuous improvement, without the necessity of another disruptive modification in the future. Furthermore, as network management is important for the reliable and safe operation of networks, it is also crucial for the success of the FI. In the scope of these challenges, the ***Self-NET Project*** (*https://www.ict-selfnet.eu/*) aims to integrate the self-management and cognition features and the inevitable part of FI evolution.

3 COGNITIVE NETWORK MANAGEMENT ACTIVITIES IN THE SELF-NET SCOPE

The ***Self-NET Project*** designs, develops and validates an innovative paradigm for cognitive self-managed elements of the Future Internet. Self-NET engineers the FI based on cognitive behaviour with a high degree of autonomy (Chochliouros et *al*., 2009b) by proposing and examining the operation of self-managed FI elements around a novel "feedback-control cycle" (i.e. the "Monitoring/Decision-Making/Execution" or "MDE" cycle) as shown in Figure 1. Thus, dynamic distribution of resources according to network needs at specific time intervals can be pursued by introducing the "MDE" cycle in order to overcome bottlenecks and ensure seamless service provisioning – *even in case of services with high bandwidth requirements*. The completion of the aforementioned objective can make certain better QoS, *beyond the original best-effort status*, and simultaneously eases operational and network management functionalities.

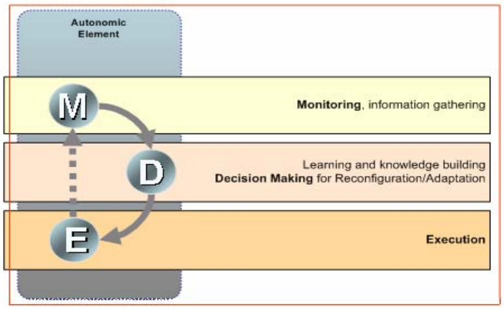


Figure 1: The generic Cognitive MDE cycle.

Cognitive management in FI elements introduces also innovative techniques regarding converged infrastructures with ultra-high capacity optical transport/access networks and converged service capability across heterogeneous environments. Furthermore, the introduction of cognition in networks can contribute towards overcoming structural limitations of current infrastructures - which render it difficult to cope with a wide variety of networked applications, business models, edge devices and infrastructures - so as to guarantee higher levels of scalability, mobility, flexibility, security, reliability and robustness. Self-NET principle design is based on high autonomy of network elements in order to allow distributed management, fast decisions, and continuous local optimization either of existing networks or of specific network parts (Kousaridas et *al*., 2008).

The three distinct phases of the Generic Cognitive Cycle Model-GCCM (i.e. the MDE cycle) are as illustrated in Figure 2:

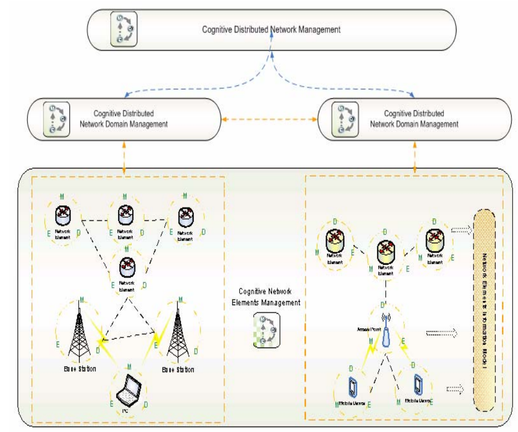


Figure 2: The proposed Distributed Cognitive Cycle for Systems and Network Management (DC-SNM).

Cognitive capabilities can enable the perception of the NEs environment and the decision upon the necessary action (e.g. configuration, healing, protection measures, etc.). As current management tasks are becoming overwhelming, Self-NET embeds new management capabilities into NEs in order to take advantage of the increasing knowledge that characterizes the daily operation of mobile FI users (Raptis et *al*., 20009). Among the main Self-NET's efforts is "*to tackle complexity*" by following the well-known "divide and conquer" approach, that is by: "*Breaking down the overall network management task into smaller manageable tasks and assigning them to individual network elements*"; showing NEs how to tackle the relevant issues; giving NEs the ability to "learn" in order to solve new, emerging (and occasionally "unforeseen") problems; facilitating NEs to cooperatively solve problems that require a sort of coordination, and; enhancing Future Internet with inherent management capabilities *(i.e. "making FI self-manageable")*. NEs with cognitive capabilities aim at fast localised decision making and (re-) configuration actions as well as learning capabilities that improve elements behaviour. An essential target of the Project effort is to develop innovative cross-layer design optimization approaches that alleviate the shortcomings and duplication of functionalities in different protocol layers of the present IP stack. Furthermore, Self-NET also provides a peer-to-peer style distribution of responsibilities among self-governed elements of the FI, therefore overcoming the barrier of current client-server and proxy-based models in the operation of mobility management, broadcast-multicast, and QoS mechanisms. A Self-Net's "key-objective" is the provision of a holistic architectural and validation framework that unifies networking operations and service facilities of the FI (Mihailovic et *al.,* 2009a). To this aim, a real-world demonstrator has been developed to "test" the applicability, robustness, and stability of all corresponding solutions and to ensure "smooth migration" from present Internet mechanisms. In parallel, simulation campaigns verify the scalability-performance sustainability of the corresponding artifacts.

FI design is required to provide answers to a number of current Internet's deficits, especially when the danger of increased complexity is more than evident. ***Self-management and autonomic capabilities*** can so alleviate this "drawback" by: Providing inherent management capabilities; increasing flexibility, and; allowing an ever-evolving Internet. Towards realizing this aim, the Self-NET Project considers that a DC-SNM along with a hierarchical distribution over the network can "map" self-management capabilities over FI architectures (Self-NET Project, 2008). DC-SNM further facilitates the promotion of distributed (or decentralized) management over a hierarchical distribution of management and (re)configuration making levels: (a) to (autonomic) network elements; (b) to network domain types, and; (c) up to the service provider realm, hence allowing high autonomy of network elements with cognitive capabilities aimed at fast localised (re)configuration actions and decision making. Such a distribution brings about the intriguing issue of orchestrating the cognitive cycles (M-D-E) at higher levels of the self-management distribution. The "*decomposition*" of network management into responsibility areas (as shown in Figure 2) can provide the principle on which universal management architecture can be developed, having as a main goal the efficient handling of complexity towards FI environments.

Such decomposition, combined with the introduction of cognitive functionalities at all layers, can allow decisions and configurations at shorter time-scales (Agoulmine et *al.*, 2006). Each element at the identified layers has embedded cognitive cycle functionalities and also the ability to manage itself and make appropriate local decisions. For an efficient and scalable network management, where various stakeholders/actors participate, a distributed approach is thus adopted. Dynamic network (re-) configuration in many cases is based on cooperative decision of various FI elements and distributed NM service components. Hints and requests-recommendations are exchanged among the corresponding layers, in order to "indicate" (or to "identify") a new situation or an action for a "targeted" execution. The automated and dynamic incorporation of various layers requirements (e.g., SLAs (service level agreements)) into the management aspects also provides novel features to NM capabilities. Moreover, the resolution of conflicting requests becomes an issue of situation awareness and elements' domain policy prioritization (Strassner, 2003). In the context of the Self-NET Project, the introduction of a hierarchical cognitive cycle to enable multi-tier self-management in various network elements and dynamic network compartments provides a quite promising approach to alleviate management overhead, ensure dynamic adaptation to service requirements, situation aware NM and reconfiguration, while coping with the fragmentation of contemporary centralised network management dedicated to specific types of networks (Elliott and Heile, 2000; Chochliouros et *al*., 2009a). Network management is a wide thematic area, including device monitoring, application management, security, ongoing maintenance, service levels, troubleshooting, planning, and other tasks – *ideally all coordinated and supervised by an experienced and reliable "entity"* (known as the "*network administrator*"). For businesses of all sizes, it is imperative to consider a NM solution that is easy to use, quick to deploy, and offers low total cost of ownership. Such a solution implicates comprehensive capabilities and a satisfactory reliability. The adoption of appropriate cognitive techniques on different platforms (and/or on parts of them, also including connected devices) can be the "kick-off" that will encourage the creation of new networking infrastructures. Furthermore, it is essential to perform NM activities in a distributed (and autonomous) way by incorporating self-organization and self-management principles, as outlined by the entire Self-NET effort (Self-NET Project, 2009). This implicates several distinct advantages and/or expected benefits, as the latter have been discussed in the scope of previous project works (Self-NET Project, 2008; 2009). Although there is a diversity of external and influencing available definition on self-management related work (Miller, 2008), the term ***"self-management"*** is applied here as the general term describing all autonomic and cognition-based operations in a system. Six relevant distinct methods are identified with specific realizations and purposes; they all serve to demonstrate principles and concepts inherent in the system properties, for achieving completeness and accuracy (Chochliouros et *al*., 2009b; Mihailovic et *al*., 2009a).

4 CHALLENGES AND BENEFITS FOR THE MARKET SECTOR

The implementation-inclusion of suitable cognitive techniques/systems on diverse platforms can be the "first step" to support the development of new networking infrastructures (Prehofer and Bettstetter, 2005). The introduced, by the Self-NET project, functionalities, can implicate major benefits for all relevant "actors", that is for both operators and users. These may include, among others:

Automatic network planning and reduction of management time of complex network parameters (and/or structures): Both present and future anticipated high proliferation of different services that a communications network should offer and support, imposes a decisive challenge for any network operator involved, while implicating an appropriate "adjustment" of network performance together with "optimization" of the network resource usage. Daily (human) network manager activities include many tedious and time-consuming tasks, to make certain that the network delivers the desired services to its users. In many cases, the network operator is obliged to search through vast amounts of monitoring data to find any "*inconveniences*" to his network behaviour and to ensure a proper delivery of services. Embedding self-management functionalities in future NEs and establishing cognition at the diverse network levels (e.g., network elements, network compartments and network domains) can automate the detection of any abnormal (or "adverse") behaviour, the remoteness of the relevant source(s), the diagnosis of the corresponding fault(s) and the expected repair of the conceived problematic situation. For a variety of reasons affecting the competitive presence of an operator in the market sector, it is a matter of high importance for the network to be able to "predict" irregular events (like faults or intrusions) and to react, *accordingly*, in due time. Thus, applying self-aware techniques in a modern network environment can ease network composition and network planning procedures and can ensure the automatic adaptation of networks/-services to capabilities of the network components.

Options for reduction of network operational cost: Any infrastructure that can perform automated operational tasks to optimize its network efficiency and the quality of service(s) offered, can contribute to the objective of reducing actual network operational expenditures (OPEX). The option for automating several procedures can be remarkably beneficial to network operators as it facilitates various complex (and resource-consuming) processes, currently deployed at a large time scale and requiring significant human intervention. This also allows for a more inexpensive and simpler network deployment: That is, by applying self-management techniques intending to optimize the network in terms of coverage, capacity, performance etc., operators can decrease their operational expenditures by limiting the manual effort required for network operation and can actively utilize their network elements (or resources) more efficiently. In addition, such techniques can simplify network maintenance and fault management.

Options for easy "network adaptation" (e.g., in new traffic models and schemes): Traffic management of a communications network is mainly based on integrated and centrally coordinated deployment of specific measures and suitable rules, in response to the current network operating state and/or in anticipation of future needs and relevant traffic conditions. Traffic management configuration of large wireless networks consisting of multiple, distributed NEs of varying technologies, is challenging, time-consuming, prone to possible errors and requires highly expensive control & management equipment from any market operator. Even when it is originally deployed, it involves continuous upgrading and modifications in order to provide a consistent and transparent service environment, to sustain high QoS, to recover from potential faults and to maximize the overall network performance, especially when congestion appears.

Seamless users' experience in dynamic network selection: In the scope of a liberalised and competitive market, the end-user desires to have access to a network that offers adequate coverage together with services of high quality, on a real-time basis. To this aim, self-management can provide decentralized monitoring and proper decision-making techniques so that appropriate optimization hints can be extracted, in terms of determining the optimum course of actions to improve network performance and stability and to guarantee service continuity.

Enhanced service provision and adaptability: The dynamic detection of operational deficiencies or poor QoS delivered to the end-user, both imply for specific remediate actions to compensate for the related problematic situation(s). Improving the overall network quality also increases subscribers' satisfaction and trust. Thus, the optimization of all procedures in order to "minimize" (or occasionally to "*delete*") service failures and to ensure continuity of service delivery is a critical issue for the user and the operator, in a competitive telecommunications market. In addition, it is quite critical for the entire network to incorporate options and/or other NM facilities able to fulfil any requirement for novel (or innovative) service features, including, *for example,* network (or service) reconfiguration capabilities, broadband management, support of an increased number of services/facilities offered.

Enabling effective networking under highly demanding conditions: A continuous and dynamically updated NM (proactively and reactively adapted to the network dynamics) is an appropriate tool for such a purpose: That is, instead of using manual techniques, a fully automated, transparent and intelligent traffic management functionality can be much more beneficial. The proposed Self-NET infrastructure can be used to provide an efficient real-time traffic management in a large network, thus maximizing network performance and radically decreasing human intervention. Several among the (potential) application areas can cover cases of traffic congestion, network attachments, link failures, performance degradation, mobility issues, multi-service delivery enhancements and involve intelligent autonomic congestion management and traffic routing, dynamic bandwidth allocation and dynamic spectrum re-allocation (Mihailovic et *al.,* 2009a).

The continuity of service availability influences directly the technical approach of service realization and is an important parameter that affects network planning; in fact, the network should possess fitting techniques to "*adapt itself*" to an essential (occasionally prescribed) functional state. To this aim, the network should be able to gather information about various entities (elements, domains, sectors) and/or distinct modules, to detect their operational state(s) and to react to any deviations from the proposed "desired" operational state. Applying self-aware mechanisms can conduct to network performance optimization, in terms of coverage and capacity, optimization of QoS delivered to the end-user, reduction of human intervention in terms of determining the most appropriate course of actions and proceeding to the implementation of optimization activities (Mihailovic et *al.,* 2009b). Consequently, applying self-aware mechanisms and contexts in future networks will contribute to guarantee some critical features including but not limited to: (i) High availability and seamless continuity of services; (ii) Connectivity *anywhere and anytime*; (iii) Robustness and stability/steadiness of the underlying network infrastructure; (iv) Scalability in terms of features and functions; (v) Optimal balance between cost network-related benefits (OPEX reduction – optimized network functionalities), and; (vi) Heterogeneity support in terms of system components and services.

5 CONCLUSIONS

Evolution towards NGNs/FI requests a more flexible and flexible architecture that will act as the "basis" for the disposal of a multiplicity of services-facilities with optimized quality levels, intending to attract and to satisfy end-users. Such modern network infrastructures are characterized by the inclusion of embedded intelligence "*per element*" or "*per domain*", targeting at a more distributed environment both in terms of management and operational activities. To this specific aim, cognitive networks with self-aware functionalities introduce a high level of autonomy, meaning that embedded and/or inherent management functionality in several components of FI systems do compose management upon a "*per NE*" and/or a "*per domain*" mechanism, rather than a centralized (traditional) network functionality. Compared to current network features, self-management techniques pave the way towards automated network processes such as the deployment of new NEs, the reconfiguration of the network (in whole or in part) and the selection/execution of the optimal corresponding solution (or "response") based on specific circumstances and remediation of identified malfunctions with the minimum potential service interruption. Consequently, new methods (related to embedded and/or autonomous management, virtualization of systems and network resources, advanced and cognitive networking of information objects), have to outline and re-define the overall FI network architecture. To "encounter" such critical challenges, the main objective of Self-NET project effort is to describe and evaluate/analyze new paradigms for the management of complex and heterogeneous network infrastructures-systems (such as cellular, wireless, fixed and IP networks), taking into consideration the next generation Internet environment and the convergence perspective. Thus, it can efficiently integrate new operational capabilities in the "*underlying system*" by introducing innovative self-management attributes, resulting in noteworthy benefits for all actors involved. The Self-NET initiative develops self-management features that alleviate consequences of events for which the system would require various invocations of remedy actions and/or significant human intervention. This dynamic behavior and intelligence of handling various events (and/or situations) can potentially lead to an innovative and much promising beneficiary scope of the entire system's operations.

ACKNOWLEDGEMENTS

The paper has been composed upon the context of the *Self-NET* ("*Self-Management of Cognitive Future Internet Elements*") European Research Project and has been supported by the Commission of the European Communities - *Information Society and Media Directorate General* (FP7-GA No.224344).

REFERENCES

Afuah, A., Tucci, C.L., 2000. *Internet Business Models and Strategies: Text and Cases,* McGraw-Hill.

Agoulmine, N., Balasubramaniam, S., Botvitch, D., Strassner, J. et *al.*, 2006. Challenges for Autonomic Network Management. In *Proceedings of the 1st IEEE International Workshop on Modelling Autonomic Communications Environments, Dublin, Ireland (October 25-26, 2006)*

Blumenthal M.S., Clark, D.D., 2001. *Rethinking the design of the Internet: The End-to-End arguments vs. The Brave New World,* ACM Transactions on Internet Technology, vol.1(1), 70-109.

Boccalettia, S. , Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U., 2006. *Complex networks: Structure and dynamics, Elsevier Physics Reports*, vol.424, 175-308.

Chochliouros, I.P., Spiliopoulou, A.S., 2003. *Innovative Horizons for Europe: The New European Telecom Framework for the Development of Modern Electronic Networks and Services*, The Journal of The Communications Network (TCN), vol.2(4), 53-62.

Chochliouros, I.P., Spiliopoulou, A.S., 2005. *Broadband Access in the European Union: An Enabler for Technical Progress, Business Renewal and Social Development,* The International Journal of Infonomics (IJI), vol.1, 5-21.

Chochliouros, I.P., Alonistioti, N., Spiliopoulou, A.S., et al., 2009a. Self-Management in Future Internet Wireless Networks: Dynamic Resource Allocation and Traffic Routing for Multi-Service Provisioning. In *Proceedings of the 1st International Conference of Mobile and Lightweight Wireless Systems - MOBILIGHT 2009, 18-20 May 2009, Athens, Greece.*

Chochliouros, I.P., Spiliopoulou, A.S., Georgiadou, E., Belesioti, et *al*., 2009b. A Model for Autonomic Network Management in the Scope of the Future Internet. In *Proceedings of the 48th FITCE International Congress, Prague, Czech Republic, pp.102-106 (September 03-05, 2009)*.

Clark, D., Sollins, K., Wroclawski, J., Katabi, et *al.*, 2003. New Arch: *Future Generation Internet Architecture (Final Technical Report)*, The US Air Force Research Laboratory.

Commission of the European Communities, 2008. Communication on "Future Networks and the Internet" [COM(2008) 594 final, 29.09.2008], Brussels.

Commission of the European Communities, 2009a. *Communication on "A public-private partnership on the Future Internet" [COM(2009)479 final, 28.10.2009]*, Brussels.

Commission of the European Communities. 2009b. *Consultation on the Future "EU 2020" Strategy" [COM(2009) 647 final, 24.11.2009*], Brussels.

Dobson, S., Denazis, S., Fernandez, A., Gaïti, et *al*., 2006. *A survey of autonomic communications.* ACM Transactions on Autonomous and Adaptive Systems (TAAS), vol.1 (2), 223-259.

Elliott, C., Heile, B., 2000. Self-organizing, self-healing wireless networks. In *Proceedings of IEEE International Conference on Personal Wireless Communications, pp.355-362 (December 17-20, 2000)*

European Future Internet Portal, 2010. [http://www.future-internet.eu/].

Future Internet Assembly (FIA), 2009. *Position Paper: Real World Internet*. [http://rwi.future-internet.eu/index.php/Position_Paper].

Galis, A., Brunner, M., Abramowitz, H, 2008 (December). *MANA Position Paper: Management and Service-aware Networking Architectures (MANA) for Future Internets. Draft 5.0*.

International Telecommunication Union - Telecommunication Standardization Sector (ITU-T), 2000. *Recommendation M. 3400: TMN Management Functions*, Geneva.

Kim, S.-S., Choi, M.-J., Ju, H.-T., Ejiri, M., Won-Ki Hong, J., 2008. Towards Management Requirements of Future Internet. In *Challenges for Next-Generation Network Operations and Service Management, pp.156-166. Berlin / Heidelberg (2008).*

Kousaridas, A., Polychronopoulos, C., Alonistioti, N., Marikar, A., et *al.*, 2008. Future Internet Elements: Cognition and Self-Management Design Issues. In *Proceedings of the 2nd International Conference on Autonomic Computing and Communication Systems (SAC-FIRE Workshop), Autonomics 2008, Turin, Italy, (September 23-25, 2008)*.

Mihailovic, A., Chochliouros, I.P., Kousaridas, A., Nguengang, G., Polychronopoulos, C., Borgel, J., Israel, et *al.*, 2009a. Architectural Principles for Synergy of Self-Management and Future internet Evolutions. In *Proceedings of the ICT Mobile Summit 2009, Santander, Spain (June 10-12, 2009)*.

Mihailovic, A., Chochliouros, I.P., Georgiadou, E., Spiliopoulou, A.S., et *al.*, 2009b. In *Proceedings of the*

International Conference on Ultra Modern Telecommunications (ICUMT-2009), *Saint Petersburg, Russia (October 12-14, 2009).*

Mihailovic, A., Nguengang, G., Kousaridas, A., Israel, M., et al., 2010. An Approach for Desiging Cognitive Self-Managed Future Internet. In *Proceedings of the Future Network & Mobile Summit 2010, Florence, Italy (June 10-12, 2010).*

Miller, B., 2008 (March). The autonomic computing edge: Can you CHOP up autonomic computing? IBM Corporation.

Organization for Economic Co-operation and Development (OECD), 2008. *The Seoul Declaration for the Future of the Internet Economy,* OECD, Paris.

Pastor-Satorras, R., Vespignani, A., 2004. *Evolution and Structure of the Internet: A Statistical Physics Approach*, Cambridge University Press, Cambridge.

Prehofer, C., Bettstetter, C., 2005 (July). *Self-Organization in Communication Networks: Principles and Design Paradigms,* IEEE Communications Magazine, vol.43 (7), 78-85.

Raptis, T., Polychronopoulos, C., Kousaridas, A., Spapis, A*., et al., 2009.* Technological Enablers of Cognition in Self-Manageable Future Internet Elements. In *Proceedings of the First International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE-2009) - Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns, 2009. COMPUTATIONWORLD'09. Athens/Glyfada, Greece pp.499-504 (November 15-20, 20090.*

Reding, V., 2009. *Internet of the future: Europe must be a key player (Speech of February 02, 2009) - Future of the Internet initiative of the Lisbon Council*, *DG Information Society and Media*, Brussels.

Self-NET Project, 2008. *Deliverable D1.1: System Deployment Scenarios and Use Cases for Cognitive Management of Future Internet Elements (October 2008)*.

Self-NET Project, 2009. *Deliverable D5.1: First Report on Business Opportunities (February 2009).*

Strassner, J., 2003. *Policy-Based Network Management*, Morgan Kaufmann Publishers, San Francisco.

Timmers, P., 1998. *Business Models for Electronic Markets*, International Journal on Electronic Markets and Business Media, vol.8(2), 3-8.

Tselentis, G., Domingue, L., Galis, A., Gavras, A., Hausheer, D., Krco, S., Lotz, V., Zahariadis, T., 2009. *Towards the Future Internet - A European Research Perspective*, IOS Press, Amsterdam.

EFFICIENT ECHOES AND PARAMETERS ESTIMATION

OFDM signals analysis in multipath scenarios

Miguel Ángel Ucha Cuevas, José Luís Ledo Gavieiro

Sistemas Integrados de Servicios de Telecontrol,S.L (GSERTEL, S.L.)., Volta do Castro s/n,15706

Santiago de Compostela,Spain

maucha@gsertel.com, jlgav@gsertel.com

Keywords: Multicarrier Modulation, channel estimation, FFT, Inter-Symbol Interference.

Abstract: This paper introduces a study of Multicarrier Modulation (MM) signals in a multipath scenario obtaining a closed expression for estimating the modulation parameters and the echoes of the channel. The method proposed is highly robust to channel impairments like noise or Inter-Symbol Interference (ISI). It is feasible to adapt it to future schemes of modulation in a really easy way and the method can cope with multi-standard MM at the same time without increasing algorithm complexity drastically.

1 INTRODUCTION

Orthogonal Frequency Division Multiplex (OFDM) has raised a lot of interest since several years ago due to its facility to combat multipath in different scenarios. Multipath impairment means that different symbols arrive at the same time to receiver, what is called Inter-Symbol Interference (ISI). To overcome this problem, OFMD insert a space between symbols. This space is called Guard Interval (GI) and its duration depends on the length of the channel. If GI is longer than channel length, ISI will be avoided.

Channel length is a difficult parameter to estimate because it usually varies on time and when we choose the length of GI we must make sure that we have room for the longest path. Channel length is related to the coherence time, which is a statistical parameter of the channel. When channel length is longest than GI ISI is present at the receiver and receiver's efficient will be highly damaged.

To estimate channel length several techniques can be used. Simulations can be helpful but they never will be accurate enough because channel usually varies on time. The most common method is to use measurement equipments to analyse the signal quality in different points through the coverage area and double-check whether the echoes are inside the GI or not. Measurement equipments demodulate the signal and use carrier pilots to carry out the channel estimation. As a result, if ISI is strong, the receiver will be unable to demodulate the signal.

To overcome those problems, an efficient echo estimation method is proposed in this paper without the need of demodulating the signal. Consequently, echoes can be estimated even when ISI is strong enough to demodulate the signal or the Signal-to-Noise-Ratio (SNR) is low. If OFDM or SC-OFDM is the modulation scheme chosen, the method proposed can be used to discover several modulation parameters like GI length, number of carriers and signal bandwidth. The method is readily adaptable to handle signals of the last generation telecommunications standards like DVB-C/C2 or DVB-T/T2 signals. One of its main advantages is that it doesn't need any side information to estimate the echoes.

All the details given in this paper are patent pending.

The method has been implemented successfully in measurement equipments with system architecture composed by a FPGA and a microprocessor. It has been tested in different scenarios showing very good performances.

2 SYSTEM MODEL

OFDM convert data bits into K parallel subcarrier with lower data rate. Inverse Fast Fourier Transform (IFFT) of N points is performed to obtain time-domain signal. After that, GI is slotted in copying P samples from the end of the symbol to the beginning. Let P be the length of the GI in samples, we know that

$$P = N \cdot G \tag{1}$$

where G (G < 1) is usually chosen as a fraction of the length of the symbol (N). Due to GI, the total length of one OFDM symbol is expressed as

$$N_s = N + P \tag{2}$$

N_s, N and P are expressed in samples not in temporal units.

Several authors (T.M. Schmidl and D.C. Cox, 1997, pp 1613-1621) (Nir, Waterschoot, Moonen and Duplicy, 2009) take profit of GI calculating an autocorrelation of the signal to detect the modulation parameters. In this paper we develop a complete different method to obtain that information.

Channel response (h(t)) can be expressed as a sum of L different paths. One of them is considered the main path and the others L-1 are secondary. So

$$h(t) = \sum_{k=0}^{L-1} \beta(k)\delta(t-\tau_k) + \mathrm{n(t)} \tag{3}$$

where $\beta(k)$ is the amplitude of the k component, τ_k its delay and $n(t)$ the noisy component of the channel. This equation can describe fundamentally two types of channels: Rice channels and Rayleigh channels. Rice channels are characterized by having a main path with amplitude higher than the others paths. On the other hand, Rayleigh channels have similar amplitudes in all paths.

Rice channel are useful to depict scenarios in which both transmitter and receiver are fixed, and signal get to the receiver following a direct path and a series of secondary paths caused by reflections of the signal in different parts of the scenario.

3 THEORICAL ANALYSIS

We will focus our study in ricean channels, where a main path is present. In such case, we will suppose, without loss of generality, that the main path arrives to receiver at t=0, so $\tau_0 = 0$ (we will consider k = 0 the main path).

3.1 The General Case

The received signal is given by

$$r(t) = h(t) * x(t) \tag{4}$$

where x(t) is the transmitted signal.
The frequency response of r(t) is given by

$$R(\omega) = H(\omega) \cdot X(\omega) \tag{5}$$

$$R(\omega) = \beta(0)X(\omega) + \sum_{k=1}^{L-1} \beta(k)X(\omega)e^{-j\omega\tau_k} \tag{6}$$

In Equation (6), the noisy component n(t) has been neglected.

If we normalize all amplitudes by $\beta(0)$, the amplitude of the main path, Equation (6) can be reformulated as

$$R(\omega) = \left(1 + \sum_{k=1}^{L-1} \alpha(k)e^{-j\omega\tau_k}\right)X(\omega) \tag{7}$$

$$\alpha(k) = \frac{\beta(k)}{\beta(0)} \tag{8}$$

The frequency response of the channel is given by

$$H(\omega) = 1 + \sum_{k} \alpha(k)e^{-j\omega\tau_k}, k \neq 0 \tag{9}$$

where, for the sake of notation, L has been omitted.

The power spectrum of the received signal is

$$|R(\omega)|^2 = |H(\omega)|^2 \cdot |X(\omega)|^2 \tag{10}$$

We analyse $R(\omega)$ in a frequency range in which the following constraint is valid

$$|X(\omega)|^2 \approx K, \qquad \omega_{inf} \leq \omega \leq \omega_{sup} \tag{11}$$

where K is a constant value. As a result, we can do the following approximation

$$|R(\omega)|^2 \approx |H(\omega)|^2 \cdot K \qquad \omega_{inf} \leq \omega \leq \omega_{sup} \tag{12}$$

The previous assumption can be done in the bandpass of digital signals, as we can see in Figure 1.

Now, we are ready to compute the square of the module of $H(\omega)$.

$$\begin{aligned} H_e(\omega) &= |H(\omega)|^2 \\ &= \left|1+\sum_k \alpha(k)e^{-j\omega\tau_k}\right|^2 \\ &= \left(1+\sum_k \alpha(k)\cos\left(\omega\tau_k\right)\right)^2 + \left(\sum_k \alpha(k)\sin(\omega\tau_k)\right)^2 \\ &= 1+\sum_k \alpha^2(k)\cos^2(\omega\tau_k) + \sum_k\sum_j \alpha(k)\alpha(j)\cos(\omega\tau_k)\cos(\omega\tau_j) + \sum_k \alpha^2(k)\sin^2(\omega\tau_k) + \sum_k\sum_j \alpha(k)\alpha(j)\sin(\omega\tau_k)\sin(\omega\tau_j) \end{aligned} \tag{13}$$

$$\forall k \neq j, k \neq 0, j \neq 0, \omega_{inf} \leq \omega \leq \omega_{sup}$$

Appling different trigonometric equalities to Equation (13) we get

$$\begin{aligned} H_e(\omega) = {} & 1+\sum_k \alpha^2(k) \\ & + \sum_k\sum_j \alpha(k)\alpha(j)\cos\left(\omega\left(\tau_k-\tau_j\right)\right) \\ & + \sum_k \alpha(k)\cos(\omega\tau_k) \\ & \forall k \neq j, k \neq 0, j \neq 0 \end{aligned} \tag{14}$$

To get a temporal expression of Equation (14) we compute the Inverse Fourier Transform (IFT).

$$\begin{aligned} h_e(t) = {} & c_0\delta(t) + \\ & \frac{1}{\pi}\sum_k\sum_j c_{kj}\delta\left(t-\left|\tau_k-\tau_j\right|\right) + \\ & \frac{1}{\pi}\sum_k c_k\delta(t-|\tau_k|) \\ & \forall t \geq 0, \forall k \neq j, k \neq 0, j \neq 0 \end{aligned} \tag{15}$$

Where

- $h_e(t)$ is the temporal response of the square of the channel frequency response.
- $c_0 = 1 + \sum_k \alpha^2(k)$ is the level of the main ray with a contribution of all remaining rays.
- $c_{kj} = \alpha(k)\alpha(j)$ is the level of the replica generated by rays k and j.
- $c_k = \alpha(k)$ is the level of ray k.

Analyzing Equation (15) the following characteristics can be distinguished:

- The main ray is located at t = 0 with a contribution of all remaining rays (level c_0).
- A replica caused by rays k and j is located at $t = \left|\tau_k-\tau_j\right|$ with the corresponding level of $\frac{1}{\pi}c_{kj}$.
- The ray k is located at $t = |\tau_k|$ with a level of $\frac{1}{\pi}c_k$.

3.2 Study of Multicarrier Signals

Now, we are ready to study the case of Multicarrier Modulations (MM). In those systems, the signal is split into blocks of N samples and a GI is inserted at the beginning of the symbol.

At the transmitter and after modulation, such as PSK or QAM, the data sequence is converted into N parallel sub-sequences, z_k^n, where n is the carrier index and k is the block index. The k^th block $z_k \stackrel{\text{def}}{=} [z_k^1, z_k^2, \ldots, z_k^N]^T$ modulates the different subcarriers computing an Inverse Fast Fourier Transform (IFFT) operation, so

$$y_k = \boldsymbol{Q}_{IFFT}\, z_k \tag{16}$$

where

$$\boldsymbol{Q}_{IFFT} = \begin{pmatrix} \frac{1}{\sqrt{N}}e^{j\frac{2\pi}{N}0} & \cdots & \frac{1}{\sqrt{N}}e^{j\frac{2\pi}{N}0(N-1)} \\ \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{N}}e^{j\frac{2\pi}{N}(N-1)0} & \cdots & \frac{1}{\sqrt{N}}e^{j\frac{2\pi}{N}(N-1)^2} \end{pmatrix} \tag{17}$$

is the IFFT transform matrix.

After that, the GI is inserted, so the time block symbol can be expressed as

$$x_k^{j+P} = y_k^{j+N} Rt\left(\frac{2j+P}{P}\right) + y_k^j Rt\left(\frac{2j-N}{N}\right) \tag{18}$$

$$\forall j \in [1-P, \dots, N]$$

where

$$Rt(n) = \begin{cases} 1, if\ |n| \le 1 \\ 0, if\ |n| > 1 \end{cases} \quad (19)$$

is the Boxcar (rectangular) function, $x_k \stackrel{\text{def}}{=} [x_k^1, x_k^2, \dots, x_k^{N+P}]^T$ is k[th] OFDM symbol after inserting GI, N is the IFFT size and P the length of GI.

The Fourier Transform of Equation (18) is given by

$$X_k(\omega)e^{j\omega T_g} = \frac{1}{2\pi}\left(Y_k(\omega)e^{j\omega T_u} * P\, sinc(\omega T_g)e^{j\omega T_g/2}\right) + \frac{1}{2\pi}\left(Y_k(\omega) * T_u\, sinc(\omega T_u)e^{-j\omega T_u/2}\right) \quad (20)$$

where * is the convolution operation and T_g and T_u are the duration of GI (P samples) and the duration of symbol (N samples) respectively.

Due to size of T_u and T_g, we can approximate $sinc(\omega T_g) \approx \delta(\omega)$ and $sinc(\omega T_u) \approx \delta(\omega)$. Bearing this approximation in mind we get

$$X_k(\omega) = \frac{1}{2\pi}\left(T_g\, Y_k(\omega)e^{j\omega T_u} + T_u\, Y_k(\omega)\right) \cdot e^{-j\omega T_g} \quad (21)$$

The first factor of Equation (21) is responsible for destroying the orthogonality of the carriers. To restore the orthogonality we must remove GI before performing the Fast Fourier Transform (FFT). The GI to Signal Ratio (GISR) is given by

$$GISR = 20\ \log_{10}\left(\frac{T_g}{T_u}\right) \quad (22)$$

As lower GISR be, less distortion will be introduced in the signal because of the GI.

The power spectrum of transmitted signal is given by

$$|X(\omega)|^2 = \frac{1}{4\pi^2}\left|T_g e^{j\omega T_u} + T_u\right|^2 |Y(\omega)|^2 \quad (23)$$

where, for the sake of notation, the sub-index k has been omitted.

Taking again Equation (10), the received signal power spectrum is given by

$$|R(\omega)|^2 = \frac{1}{4\pi^2}|H(\omega)|^2\left|T_g e^{j\omega T_u} + T_u\right|^2|Y(\omega)|^2 \quad (24)$$

We analyse $|R(\omega)|^2$ in a frequency range in which the following constraint is valid

$$|Y(\omega)|^2 \approx K, \qquad \omega_{inf} \le \omega \le \omega_{sup} \quad (25)$$

where K is a constant value. As a result, we can do the following approximation

$$|R(\omega)|^2 \approx |H(\omega)|^2 \cdot \left|T_g e^{j\omega T_u} + T_u\right|^2 \cdot K \qquad \omega_{inf} \le \omega \le \omega_{sup} \quad (26)$$

Again, the previous assumption can be done in the band pass of the MM signal, as we can see in Figure 1. The $\frac{1}{4\pi^2}$ factor has been included in K.

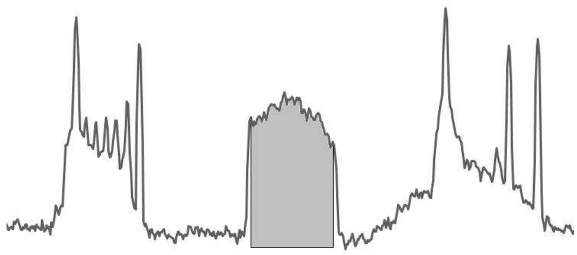


Figure 1: Frequency range used in the analysis.

Now, we define

$$H_e(\omega) = |H(\omega)|^2 \cdot \left|T_g e^{j\omega T_u} + T_u\right|^2 = \left|\left(1 + \sum_k \alpha(k)e^{-j\omega\tau_k}\right)\left(T_g e^{j\omega T_u} + T_u\right)\right|^2 \qquad \omega_{inf} \le \omega \le \omega_{sup} \quad (27)$$

After some mathematics we can solve Equation (27) and get

$$H_e(\omega) = \left(1 + \sum_k \alpha_k^2\right)\left(\mathrm{T_u^2} + \mathrm{T_g^2}\right) + 2T_u^2 \sum_k \alpha_k \cos(\omega\tau_k) + T_u^2 \sum_k \sum_j \alpha_k\alpha_j \cos\left(\omega(\tau_k - \tau_j)\right) \quad (28)$$

$$+\left(1+\sum_k \alpha_k^2\right) 2T_uT_g\cos(\omega T_u)$$
$$+4T_uT_g\sum_k \alpha_k \cos\big(\omega(\tau_k - T_u)\big)$$
$$+2T_uT_g\sum_k\sum_j \alpha_k\alpha_j \cos\,(\omega(\,\tau_k-\tau_j - T_u\,))$$
$$+2T_g^2\sum_k \alpha_k\cos\big(\,\omega(\tau_k-2T_u)\big)$$
$$+T_g^2\sum_k\sum_j \alpha_k\alpha_j\cos\Big(\omega\big(\tau_k - \tau_j - 2T_u\big)\Big)$$

$$\omega_{inf} \le \omega \le \omega_{sup}\,, \forall k \ne j, k \ne 0, j \ne 0$$

where, for the sake of notation, $\alpha_k = \alpha(k)$.

To get a temporal expression of Equation (28) we compute the IFT.

$$\begin{aligned} h_e\,(t) =& \\ &\big(T_u^2 + T_g^2\big)c_0\delta(t) \\ &+\frac{2}{\pi}T_u^2\sum_k c_k\delta(t-|\tau_k|) \\ &+\frac{1}{\pi}T_u^2\sum_k\sum_j c_{kj}\,\delta(t-|\tau_k-\tau_j|) \\ &+\frac{2}{\pi}T_uT_gc_0\delta(t-T_u) \\ &+\frac{4}{\pi}T_uT_g\sum_k c_k\delta(t-|\tau_k - T_u|) \\ &+\frac{2}{\pi}T_uT_g\sum_k\sum_j c_{kj}\,\delta(t-|\tau_k-\tau_j-T_u|) \\ &+\frac{2}{\pi}T_g^2\sum_k c_k\delta(t-|\tau_k-2T_u|) \\ &+\frac{T_g^2}{\pi}\sum_k\sum_j c_{kj}\,\delta\big(t-|\tau_k-\tau_j-2T_u|\big) \\ &t\ge 0\,, \forall k\ne j, k\ne 0, j\ne 0 \end{aligned} \tag{29}$$

where

- $h_e(t)$ is the echo function we are looking for.
- $c_0 = 1+\sum_k \alpha^2(k)$
- $c_{kj} = \alpha(k)\alpha(j)$
- $c_k = \alpha(k)$

Analyzing Equation (29) we get the following results:

- The main ray is located at t=0 with a gain of $\big(T_u^2 + T_g^2\big)c_0$.
- The ray k is located at $t = |\tau_k|$ with a gain factor of $\frac{2}{\pi}T_u^2c_k$.
- A replica of rays k and j is located at $t = |\tau_k - \tau_j|$ with a gain factor of $\frac{1}{\pi}T_u^2c_{kj}$.
- A spurious is generated at $t = T_u$ with a gain factor of $\frac{2}{\pi}T_uT_gc_0$ due to the insertion of the Guard Interval.
- A replica of ray k is generated at $t = |\tau_k - T_u|$ with a level of $\frac{4}{\pi}T_uT_gc_k$ due to GI.
- A replica of rays k and j is generated at $t = |\tau_k - \tau_j - T_u|$ with a gain factor of $\frac{2}{\pi}T_uT_gc_{kj}$ due to the presence of GI.
- A replica of ray k is generated at $t = |\tau_k - 2T_u|$ with a gain factor of $\frac{2}{\pi}T_g^2c_k$ caused by the existence of GI.
- A replica of rays k and j is generated at $t = |\tau_k - \tau_j - 2T_u|$ with a gain factor of $\frac{1}{\pi}T_g^2c_{kj}$

Bearing those results in mind, we can define a procedure to get the parameters used in the MM and the echo channel response.

Usually we are only interested in $0 \le t \le T_u$, so all factors of Equation (29) which are outside this interval can be ruled out.

The first step after getting Equation (29) is to normalize it by the gain of the main ray (located at t=0). After that, if we know all possible values of T_u, which are usually specified in the standard, we can look for the spurious that is supposed to be generated at that position. In that way we can readily find out T_u. Notice T_u is related to bandwidth of the signal and the number of carriers used in the modulation. So, the bandwidth of the signal and the number of carriers can be discovered analyzing T_u. Table 1 gives us an example of that relationship in the DVB-T system (ETSI, 2004).

To be precise, we know that $T_u \in (T_u^1, T_u^2, \cdots, T_u^K), \forall K$. Thus,

$$\begin{array}{ccc} \gamma_1 & = & h_e(T_u^1) \\ \vdots & \vdots & \vdots \\ \gamma_K & = & h_e(T_u^K) \end{array} \tag{30}$$

Now, we choose T_u as

$$T_u = \max_t\{\gamma_1,\cdots,\gamma_K\},\;\; 1\le t\le K \tag{31}$$

Table 1: Relationship among bandwidth, number of carriers and T_u in DVB-T.

	8 MHz	7 MHz	6 MHz
8K	$896\mu s$	$1024\mu s$	$1194.667\mu s$
2K	$224\mu s$	$256\mu s$	$298.6667\mu s$

The level of the spurious generated at $t = T_u$ gives us the length of the GI used in the modulation. Notice that level is dependent on T_g, so if we define a threshold for each possible GI we can easily find out the length of the GI used. So, we define N thresholds $U \in (U_1, \cdots, U_N), \forall N$ which are expected for the N possible GI. Then, we estimate GI length as

$$GI = \min_t\{|\gamma_u - U_1|, \cdots, |\gamma_u - U_N|\}, \quad 1 \le t \le N \tag{32}$$

$$\gamma_u = h_e(T_u) \tag{33}$$

The theoretic value of each threshold is calculated as follows.

We know that GI length is proportional to the length of the symbol T_u, as expressed in Equation (1), so the normalized level of the delta located at $t = T_u$ can be computed as

$$U_G = \frac{2T_gT_u}{\pi\left(T_g^2 + T_u^2\right)} = \frac{2G}{\pi(G^2 + 1)} \tag{34}$$

where G cover all possible GI used in the standard.

This threshold can be expressed in logarithmic units as

$$U_G[dB] = 20 \cdot \log U_G \tag{35}$$

Table 2: Theoretic thresholds for different GI.

G	$U_G[dB]$
$1/4$	-16.4902
$1/8$	-22.1189
$1/16$	-28.0387
$1/32$	-34.0339

Table 2 shows the theoretic thresholds for different GI used in DVB-T standard. Notice these values are complete independent on all possible echoes present in the channel.

4 EXPERIMENTAL RESULTS

The previous procedure has been implemented in a system composed by a FPGA and a microprocessor. The FPGA is responsible for getting the spectrum doing a FFT, cutting it out to analyze just the bandwidth needed, and performing the IFFT of the power spectrum obtained.

At this point we should be aware that the temporal axis t is related to the number of points of the FFT (N_{FFT}), the number of points of the IFFT (N_{IFFT}) and the frequency sampling (F_s) in the following way:

$$t_k = k \cdot \frac{N_{FFT}}{N_{IFFT}F_s}, 0 \le k < N_{IFFT} \tag{36}$$

The microprocessor is responsible for obtaining the modulation information examining the FPGA output. Moreover, it performs an algorithm which computes the level of each ray and gets rid of all replicas of the different rays. The description of that algorithm is out of the scope of this paper.

Figure 2 shows the result applying the proposed algorithm to a DVB-T signal. The modulation parameters are:

- Bandwidth: 8 MHz.
- Number of carriers: 8K.
- Guard Interval: 1/4

In this example, two echoes are present. Their position are 100 μs and 170 μs. Their amplitudes are both -5 dB below the main path amplitude. Notice the amplitudes plotted in Figure 2 are scaled by the main path amplitude and the extra factor $\left(T_u^2 + T_g^2\right)c_0$. The delta at T_u is used to detect the number or carriers, the bandwidth of the signal and the GI length. All others deltas are replicas of both echoes and must be removed to plot the real echoes.

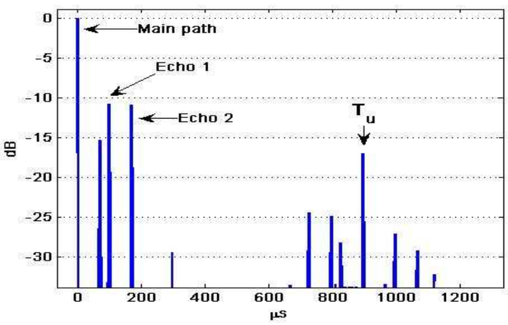


Figure 2: Echoes and parameters detection. Two echoes are present.

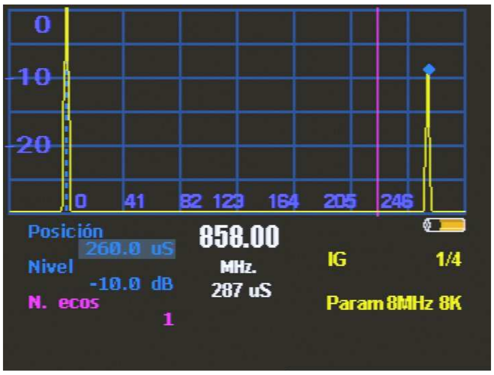


Figure 3: Echo and parameters detection. The echo is outside the GI. The GI length is 224 µs.

Figure 3 shows an echo at 260 μs (outside the GI) with a level of -10 dB. The parameters of the modulation are obtained as previously explained. In this example we can see the power of the system which is able to detect echoes falling outside the GI. Echoes outside GI produce ISI and most of the demodulators can't cope with them. In this figure, the FPGA output has been processed to get rid of all replicas generated by the echoes.

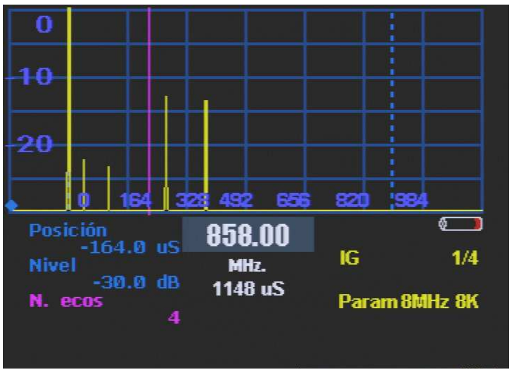


Figure 4: DVB-T signal with four echoes. Parameters: Mode 8K, GI ¼, Bandwidth 8 MHz.

Figure 4 shows a scenario in which two echoes are outside GI and the other two are inside GI. Again, the FPGA output has been processed to show the real echoes and their real levels.

5 CONCLUSIONS

We have developed an algorithm to estimate the modulation parameters used in MM systems and the echoes of the channel. In Single Carrier (SC) modulations, the algorithm can be used to obtain the echoes. The fact that demodulation is not needed makes the algorithm really robust against channel impairments. The main drawback is that it is impossible to distinguish pre-echoes and post-echoes.

REFERENCES

ETSI. (2004). *Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television.* European Telecommunication Standard. ETSI EN 300 744 v1.5.1.

T.M. Schmidl and D.C. Cox. (1997). Robust Frequency and Timing Synchronization for OFDM. *IEEE Trans. Communication*, 1997,vol. 45, pp. 1613-1621.

Vicent Le Nir, Toon van Waterschoot, Marc Moonen and Jonathan Duplicy. (2009). Blind CP-OFDM and ZP-OFDM Parameter Estimation in Frequency Selective Channels. *EURASIP Journal on Wireless Communications and Networking.* 2009, Vol. 2009, Article ID 315765. Hindawi Publishing Corporation.

AN INTEGRATED INFORMATION AND TELECOMMUNICATION TOOL IMPROVING COMPLEX TECHNOLOGIES TEACHING AND LEARNING

pilar.fernandez@ehu.es

P. Fernández-Sánchez , A. Salaverría Garnacho
Dpt. of Electronic and Telecommunication-UPV and Institute for Applied Electronics-UV

University of Pais Vasco, 20018 San Sebastián, Spain

E. Mandado Perez, Dpt. of Electronic Technology and Institute for Applied Electronics
A. Trabado
Noventia Ingenieros and Institute for Applied Electronics
Universidad de Vigo ,36210 Vigo, Spain

Keywords: integrated learning system; computer aided learning; virtual laboratory; self assessment tool.

Abstract:***This paper describes SIAPE an integrated computer-based learning system improving engineering teaching and learning of complex technologies. It combines a tutorial with a virtual laboratory and a self assessment tool to achieve a self-learning system. The tutorial is a hypermedia document linking different concepts of Electronics technology and combining them with a virtual laboratory. This virtual laboratory is made up of a set of virtual experiments with a user-friendly graphic interface and interactive simulated electronic instruments relating practical and theoretical concepts. The self assessment tool combines a test with each experiment of the virtual laboratory. The whole system matches the constructivist theory and constitutes an interactive computer-based complex tool.***

Introduction

Engineering teaching and learning has turned very difficult because present technological solutions including many interrelated concepts tend to become very complex [1] [2]. Consequently some engineering education experts [3] [4] claim that it is necessary to develop new education methods using information technology tools to improve teaching and learning. Using these methods it is possible to combine theoretical and practical activities [5] [6] in order to learn technology synthesis methods using actual products as working examples [7]. There is also a general agreement about the need of new teaching and assessment strategies to enhance competence-based learning [8]. The whole system matches the constructivist theory [9] [10] [11] [12] and constitutes an interactive computer-based complex tool enhancing competence based learning.

At present, complex technologies learning comprises two steps. First, the students learn theoretical concepts attending lectures and studying a tutorial book and second, they perform a set of tasks in a laboratory. However, this process has several drawbacks:

- Tutorials included in a book do not give enough information about the dynamic behavior of the different types of circuits and systems.
- Students have insufficient knowledge about commercial components, assembly techniques and measurement instruments. Assembly errors and component damage are common during laboratory classes.
- Students do not know if their knowledge about theoretical concepts is good enough.
- Students have no experience on testing different components behavior.
- It is not suitable for asynchronous distance learning

This paper describes a computer-based system improving the engineering education process. This system is an integrated information tool because it combines three subsystems (see figure 1):

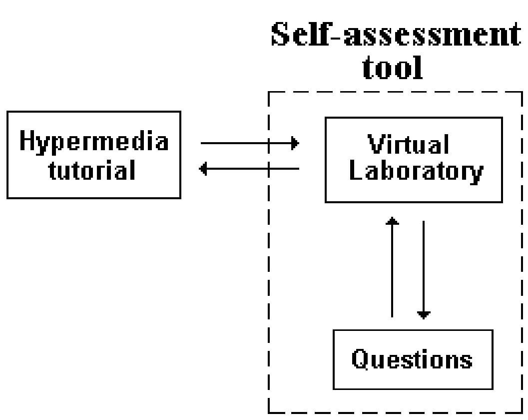


Figure 1. Hypermedia integrated learning system block diagram

Ttutorial

It is a hypermedia learning system structured into chapters, sections and subsections. The different subsections are associated with one or more virtual experiments through a user-friendly graphic interface and some interactive simulated instruments relating practical outcomes with theoretical concepts. The student can execute each experiment at the point indicated in the tutorial after studying the corresponding theoretical concept.

Virtual Laboratory

There are several kinds of virtual laboratories with different application purposes. The virtual laboratory of the integrated information tool comprises a set of virtual experiments. Each experiment is a pedagogical interactive simulation using just a few computer resources and can be integrated with some other software application.

Self-assessment tool

This is a computer tool combining one or more tests with each experiment of the virtual laboratory. In each case the student selects what he thinks is the correct answer and the system does not provide the solution but shows him the experiment outcomes, giving the possibility of verifying his answer. In this way he can learn from his mistakes [13].

As an example an "Electronics Integrated Learning Information tool" has been developed including a virtual

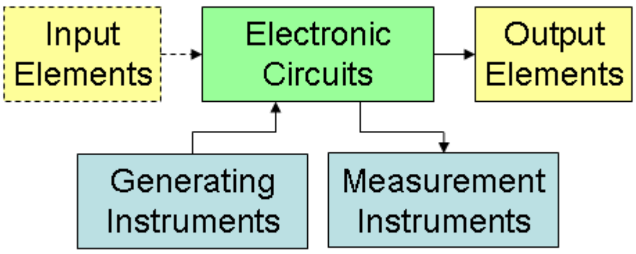


Figure 2. Block diagram for of the interface of the virtual laboratory of SIAPE

laboratory and a self assessment tool.

Virtual Laboratory

The Electronics Virtual Laboratory has the following characteristics:

- A user-friendly graphic interface.
- Interactive simulated instruments with a functionality similar to that of the actual instruments.
- Relates practical with theoretical concepts through a selected set of experiments.
- Each experiment is a pedagogical interactive simulation using just a few computer resources and can be integrated with some other software application.
- Provides destructive experiments which are not possible when using actual electronic devices.
- It has self-test capacity.
- It is scalable to add new performances.

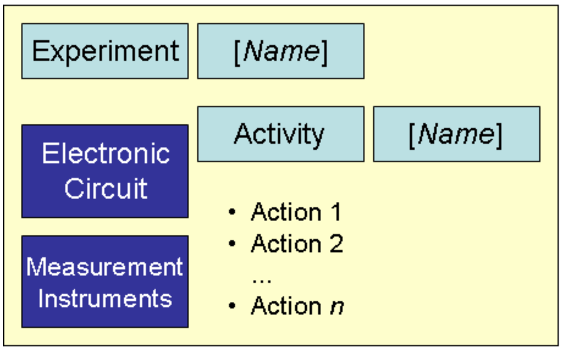


Figure 3. Block diagram of the experiment user interface

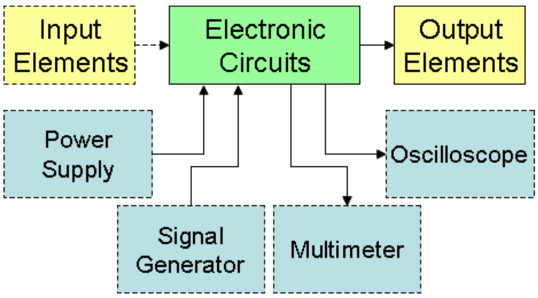


Figure 4. User interface block diagram of the Analog Electronics Virtual Laboratory

The user interface includes the different instruments used in the actual laboratory. These instruments are connected to some suitable test points. Figure 2 shows the block diagram for the user interface of the electronics virtual laboratory, including generating instruments, measurement instruments, input and output elements, and the electronic circuits under test.

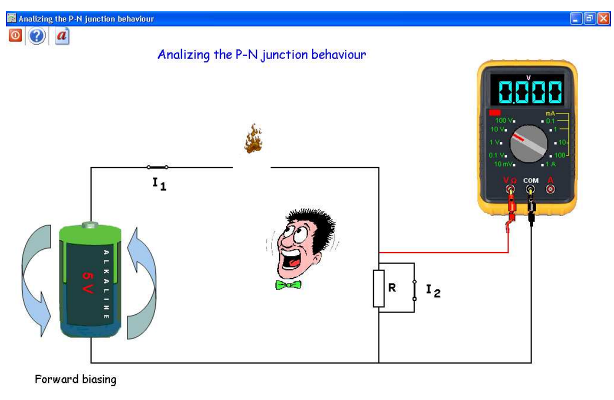


Figure 5. PN junction destructive experiment behaviour using a multimeter

Each experiment includes the description of the activities to be accomplished by the student, as shown in figure 3.

The main parts of Electronics are Analog and Digital Electronics, and a Virtual Laboratory has been developed for each one of them.

A. Analog Electronics Virtual Llaboratory

Figure 4 shows the user interface block diagram of the analog electronics virtual laboratory experiments including a signal generator, a power supply, a multimeter and an oscilloscope. All experiments use a power supply, and depending on the characteristics of the experiment, one or more of the other three available experiments.

.

The multimeter is used in the experiments where only DC signals are involved. As an example, in figure 5 shows the experiment analyzing the behaviour of the PN

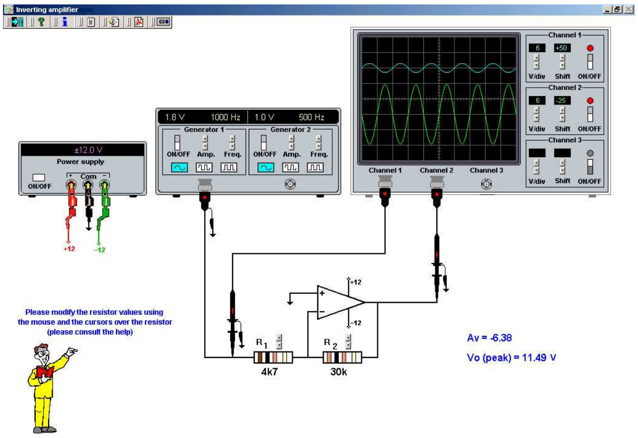


Figure.6 Inverting amplifier experiment using a signal generator and an oscilloscope

junction. This is a destructive experiment where the user can verify what happens when the PN junction is directly biased without a current limiting resistor.

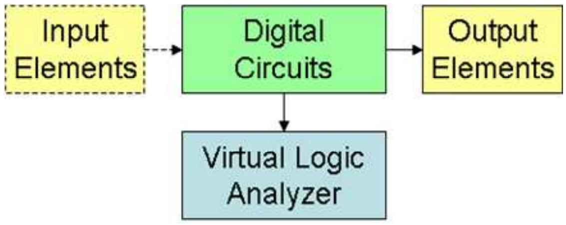


Figure 7. Digital circuits virtual experiment block diagram

The oscilloscope is used when the experiment includes time varying signals. The student can select the type of waveform, the frequency and amplitude of the signal generator output and the scope main parameters. He can also change the resistor standard values using the mouse. As an example, figure 6 illustrates the experiment analyzing the behaviour of an inverting amplifier.

Digital Electronics Virtual Laboratory

The digital electronics virtual laboratory (see figure 7) has an interactive graphic user interface including a digital circuit and a virtual logic analyzer to visualize the evolution of the different input and output signals.

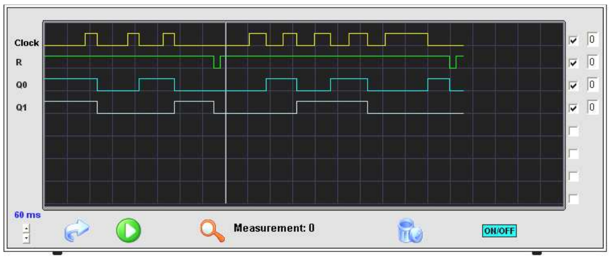


Figure 8. Virtual logic analyzer interface

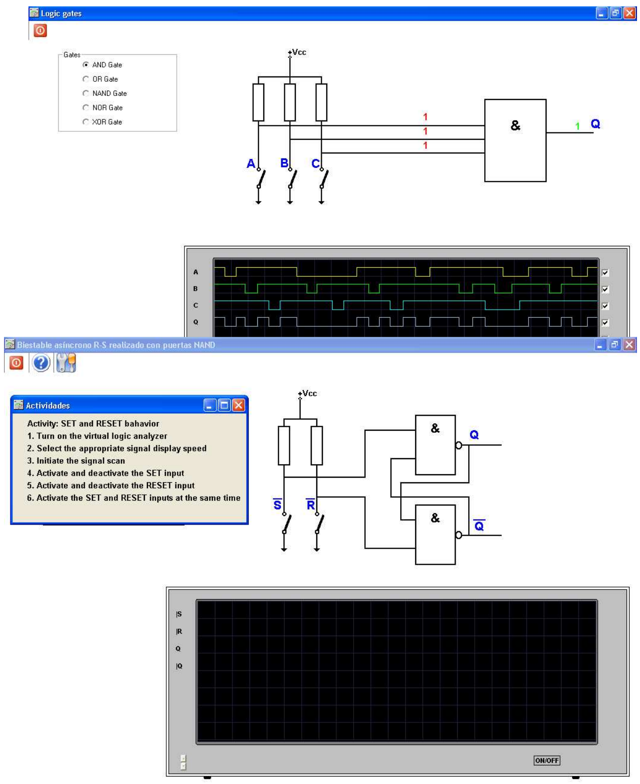


Figure 10. Student interface of the experiment “R-S latch with set priority”

The virtual logic analyzer (see figure 8) has 8 channels. It works like an actual logic analyzer, including user programmable signal display speed. Signal display can be stopped at any time. In this situation, when the user clicks on the magnifying glass a cursor facilitates signal level reading on

any point of the virtual logic analyzer screen. The virtual logic analyzer is a suitable tool to analyze the behaviour of the different combinational and sequential circuits. Figure 9 shows the experiment describing the behaviour of the different logic gates. The student can select the type of gate and the logic level of the inputs. Figure 10 shows the R-S latch behavior including its activity description.

Self Assessment Tool

The analog and digital virtual experiments can be used by the professor in the classroom but to become a useful tool improving the user self-learning capacity they must be combined with an assessment tool.

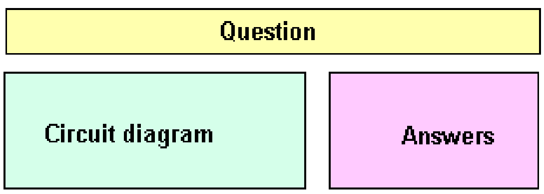


Figure 11. Assessment tool user interface block diagram

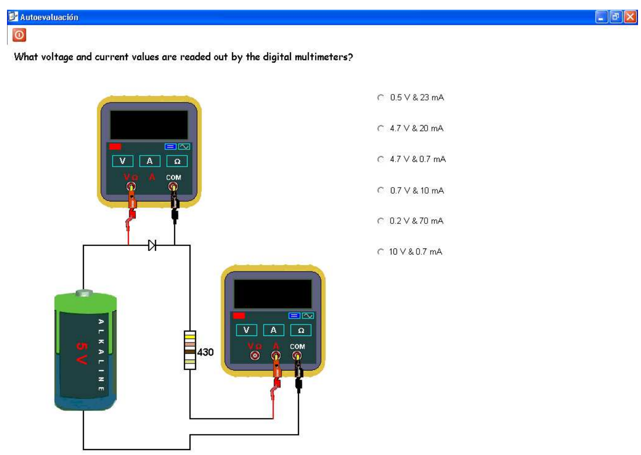


Figure 11. Example of the user interface of a self assessment question

Figure 12. Eecuting the experiment after answering the question

As shown in figure 1, after studying the hypermedia learning system combining the tutorial and the virtual laboratory, the student executes the self assessment tool. This tool is a set of multiple choice questions. Figure 11 shows the self-assessment tool user interface block diagram combining one experiment schematic, one question and the list of the possible answers. The student must select the answer that he thinks is correct. The tool does not give him the correct answer but it presents the experiment asking him to execute it to verify if the answer is correct. In this way the student has the opportunity of executing the experiment before giving the definitive answer. This process constitutes a double assessment using not only the question but also the experiment.

Figure 11 shows, as an example, one question about the semiconductor diode. Answering this question the student demonstrates his expertise about the semiconductor diode behaviour when it is forward biased. The student must choose one of the possible answers and then the tool executes the correspondent experiming. In this way the student can make voltage and current measurements again to verify if his answer is correct in the same way as in an actual laboratory (see figure 12).

Conclusions

The integrated learning system described in this paper is an enhanced active learning rich environment [11] increasing the self learning capacity of students. The virtual laboratory is structured as a set of learning objects and it can be combined with different tutorials. The self assessment tool can also be integrated with some interactive learning management system such as for example Moodle.

References

[1] Brockman J.M. *Complex Systems and Emergent Technologies*. Report of the Center for Integrated Design Seminar. June 29, 1998.

[2] Valdés M. D., Moure M. J., Mandado E.. *Hypermedia: A Tool for Teaching Complex Technologies*. IEEE Transactions on Education. Volume 42. Number 4. November 1999.

[3] Salaverría A. *Nueva metodología para la enseñanza asistida por ordenador de la Electrónica Aplicada.*Tesis doctoral. Universidad del País Vasco. 2003.

[4] Bucciarelli L. L. *Educating the Learning Practitioner*, Invited Lecturer, SEFI Annual Conference. Viena. 1996.

[5] Bucciarelli L. L. *Engineering Philosophy*. Delft University Press. The Netherlands. 2003.

[6] Felder R. M & Silverman L. K. *Learning and Teaching Styles in Engineering Education*, Engineering Education, 78 (7), 674.1988.

[7] Felder R. M. et al. *The future of engineering education II. Teaching methods that work.* Chemical Engineering Education, 34, pp. 26-39. 2000

[8] Benlloch-Dualde J. V. & Blanc-Clavero, S. (2007). *Adapting teaching and assessment strategies to enhance competence-based learning in the framework of the European convergence process*. 37th Annual *Frontiers in Education Conference.* Piscataway, NJ, USA: IEEE pp. 1-6. 2007.

[9] Tam M. *Constructivism, Instructional Design, and Technology: Implications for transforming distance learning***.** Teaching and Learning Centre. Lingnan University, Tuen Mun. Hong Kong, China. 2000.

[10] Scardamalia M., Bereiter C., McLean R. S., Swallow J. & Woodruff E. *Computer-supported learning environments and problem-solving.* Berlin: Springer-Verlag. 1989.

[11] R.S. Grabinger & J.C. Dunlap. *Rich environments for active learning: A definition.* In D. Squires, G. Conole & G. Jacobs (eds). *The Changing Face of Learning Technology.* University of Wales Press. Cardiff. 2000.

[12] Karagiorgi Y. & Symeou L. *Translating Constructivism into Instructional Design: Potential and Limitations.* Educational Technology & Society, 8 (1), 17-27. 2005.

[13] Przesmycki H. Pédagogie différenciée. Paris. Hachette. 1991.

VEHICULAR NETWORKS, TECHNOLOGIES AND CHALLENGES

Part-Escrivá, M. C., Reig-Pascual, J., Rubio-Arjona, L.
Instituto de Telecomunicaciones y Aplicaciones Multimedia, Universidad Politécnica de Valencia, Camino Vera s/n, 46022 Valencia, España
xpart@dcom.upv.es, jreigp@dcom.upv.es, lrubio@dcom.upv.es

Keywords: Vehicular communications, vehicular networks, Vehicular Ad hoc NETwork (VANET), Dedicated Short-Range Communications (DSRC), Wireless Access in Vehicular Environments (WAVE).

Abstract: Vehicular communications are more and more a focus of leading world research, driven by navigation safety requirements and by the investments of car manufacturers and Governments. The promise of safer roads and enhanced driver experience enabled by radio communication of location and traffic information in vehicular networks, has attracted effort and investment over the last few years. Vehicles have recently incorporated onboard computers, sensors, microcontrollers, etc. and the next step is to add communications among vehicles, as well as between vehicles and roadside infrastructure. New standards in vehicular networks field are also relevant in order to achieve a world market. The development of the IEEE 802.11p standard for the PHY and MAC communication layers and the IEEE 1609 suite of standards defining higher-layer functionalities, such as resource management and security, and the allocation of spectrum in the Dedicated Short-Range Communications (DSRC) band in several countries, have led to a promising new world in vehicular communications.

1 INTRODUCTION

Over the last few years, significant research efforts in industry, academia and government entities have been underway to develop Intelligent Transportation Systems (ITS). The main challenge lies in integrating wireless communications, computing and advanced sensor technologies into vehicular and transportation systems.

Road safety constitutes a major field of research, the main thrust of which is focused on reducing accident rates on the roads. For instance, in 2008 in Spain, 93,161 traffic accidents involving injuries were recorded, with 3,100 fatal, 16,488 seriously injured and 114,459 slightly injured. In the E.U. (27 members) in the same year, there were 39,051 deaths due to traffic accidents, with Spain making up 8% of the total [1]. These figures entail very high economic costs, about 2% of the EU's Gross Domestic Product, and dwarfed by the social impact, which is hard to evaluate.

Besides the obvious advantages of reducing accident rates and increasing transportation efficiency, new technologies in ITS systems also contribute to the reduction of the environmental impact of transportation systems. For example, reducing the number of accidents can in turn reduce the number of traffic jams, which could reduce the impact of CO_2 emissions into the environment.

Vehicles already have sophisticated computers and sensors onboard in order to increase effective driver assistance and protection mechanisms. For instance, passive safety mechanisms like anti-lock braking systems, navigation systems, parking radars, temperature and raining sensors, cameras, etc. The enormous difference between previous systems and new ITS systems is the addition of wireless vehicular communications, which lead to new computing and sensing capabilities.

In a vehicular network, vehicles are equipped with communication devices which allow them to collect information about traffic and road elements and exchange this information in real time with other vehicles. In addition, vehicles can also cooperate with infrastructure. Vehicles and infrastructure constitute, in this way, a decentralized self-organized cooperative ad hoc network called VANET (Vehicular Ad hoc NETwork) for road safety and transportation efficiency, as we can see in the figure 1.

It is the beginning of a new kind of ad hoc network which will hit streets: vehicles communicate with each other and with roadside infrastructure to provide a long list of innovative applications varying from transit safety to driver assistance and Internet services.

There is no single main node to manage the VANET; vehicles and infrastructure exchange messages in a self-organized way between devices on board the vehicles called OBU (On Board Unit) and devices along the roadside called RSU (Road Site Units).

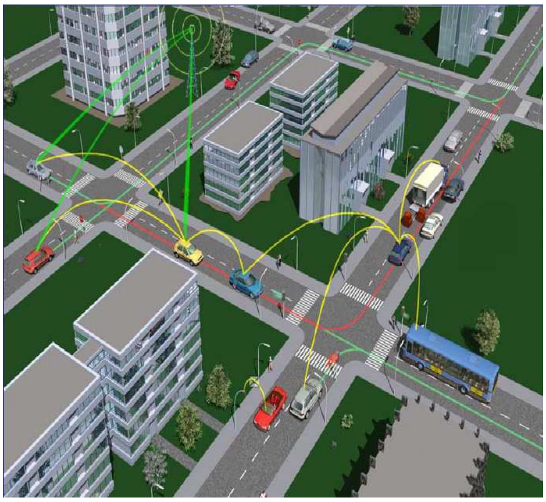
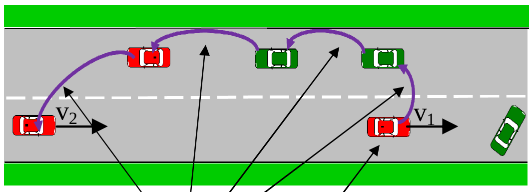


Figure 1: Information Exchange in ITS Systems (from Car to Car Communication Consortium).

Transmission of information can occur in a direct way between nearby nodes or through distant vehicles using the multihop mechanism as we can see in figure 2, where the first vehicle – running at v_1 speed- finds an obstacle and alerts the following vehicle – running at v_2 speed- using not direct transmission but communication through vehicles moving in the opposite sense. Obviously, knowledge of the real time position of nodes is necessary for successful message routing. Real time updated position information has to be reached by localization techniques which overcome GPS limitations, such as Dead Reckoning, cellular localization and image and video techniques of location.

The information flow raises the problem of security – consider, for example, the extremely serious danger of an attacker who contaminates the network with false information- and privacy – the network knows vehicle position and movements, to cite a few-. Readers can address the European SeVeCom (Secure Vehicle Communication) Projects to deal with this emergent challenge in depth. These European research projects deals with how to keep car-to-car data transmissions private and secure from malicious hackers.



Multihop transmission from first vehicle using nearby vehicles (moving in the opposite sense).

The first vehicle finds the obstacle, activates the braking system and transmits an alert message to following vehicles.

Figure 2: Multihop in vehicular network.

2 STANDARDS AND SPECTRUM

2.1 Standards

The IEEE 802.11p standard specifies the technology suitable for vehicular communication networks. This standard is an amendment to the IEEE 802.11 standard which defines the lower layers of the OSI (Open System Interconnection) model adapted for vehicular networks: PHYsical layer (PHY), Medium Acces Control layer (MAC) and Logical Link Control layer (LLC).

Within this amendment, a new operational mode, called the WAVE (Wireless Access in Vehicular Environments) mode, is defined to enable communication among high-speed vehicles or between a vehicle and a stationary roadside infrastructure network.

In addition, the IEEE 1609 family of standards for WAVE defines the architecture, communications model, management structure, security mechanisms and physical access for high speed (up to 27 Mb/s) short range (up to 1000m) low latency wireless communications in the vehicular environment. In other words, the IEEE 1609 standard suite is defined for the upper OSI layers, this is, for resource management, security services, networking services and multi-channel operation in the WAVE mode.

Primary architecture components in WAVE are OBU, RSU, and WAVE interface. They are specified in the IEEE 1609 family of standards:

1609.1: Resource Manager

1609.2: Security Services for Applications and Management Messages
1609.3: Networking Services
1609.4: Multi-Channel Operations

2.2 Spectrum

In 1999, the U. S. Federal Communication Commission allocated 75 MHz of Dedicated Short-Range Communication (DSRC) spectrum at 5.9 GHz to be used exclusively for vehicle-to-vehicle (V2V) and infrastructure-to-vehicle (V2I) communications.

The DSRC spectrum is divided into seven 10MHz wide channels as we can see in the table 1. Channel 178 is the control channel, which is generally restricted to safety communications only. The two channels at the edges of the spectrum are reserved for future advanced accident avoidance applications and high powered public safety usages. The rest are service channels and are available for both safety and non-safety usage.

Table 1: DSRC channels.

Channel number	Central frequency (MHz)	Maximum band width (MHz)
172	5860	10
174	5870	10
175*	5875	20
176	5880	10
178**	5890	10
180	5900	10
181*	5905	20
182	5910	10
184	5920	10

* Channels 175 and 181 come from the possibility of combining channels 174 and 176 or 180 and 182 in order to obtain a 20 MHz channel.

** Channel 178 is the control channel.

The spectrum use requirements for European ITS systems have been discussed in congress at the European Telecommunications Standards Institute (ETSI) for critical safety applications [2] and for all types of applications [3].

Spectrum usage as proposed in the ETSI documents is no envisaged to be on a national basis but rather on a pan European basis in order to avoid coordination/protection zone issues along borders. It also favours the usage by vehicles that travel cross border and operate in countries other than their home country.

3 CHANNEL PROPAGATION

It is widely understood that radio propagation has a significant impact on the performance of wireless communication systems. The radio channel impact on vehicular networks is even more important due to high data rate, mobility, critical situations, etc. Because of this, an accurate radio channel model has to be developed to evaluate vehicular networks in an appropriate way.

The model must reflect the three radio propagation components: path loss, shadowing and multipath fading. Path loss represents the average received signal power relative to the transmitted power as a function of the distance between transmitter and receiver. Shadowing reflects the variation of the received power at certain distance. It is a log-normal random variable. Finally, multipath fading effect results from the reception of multiple replicas of the signal.

3.1 Vehicular to vehicular radio channel model

Some relevant works set a three slope propagation model for V2V communications path loss, L:

$$L[\mathrm{dB}]=\begin{cases} 20\cdot\log_{10}\left(\frac{4\cdot\pi\cdot d}{\lambda}\right), \text{ si } d\leq d_0 \\ 20\cdot\log_{10}\left(\frac{4\cdot\pi\cdot d_0}{\lambda}\right)+10\cdot\gamma_0\cdot\log_{10}\left(\frac{d}{d_0}\right), \text{ si } d_0\leq d\leq d_1 \\ 20\cdot\log_{10}\left(\frac{4\cdot\pi\cdot d_0}{\lambda}\right)+10\cdot\gamma_0\cdot\log_{10}\left(\frac{d_1}{d_0}\right)+10\cdot\gamma_1\cdot\log_{10}\left(\frac{d}{d_1}\right), \text{ si } d>d_1 \end{cases} \quad (1)$$

where d is the distance between the transmitter and the receiver in meters, d_0 is the first breakpoint distance in meters, d_1 is the second breakpoint distance in meters, γ_0 is the path loss factor beyond the first breakpoint, γ_1 is the path loss factor beyond the second breakpoint, and λ is the carrier wavelength in meters.

The three slope propagation model is proposed in the ECC (Electronic Communications Committee) 101 Report [4], using the parameters shown in table 2, which includes a column for the two slope propagation model suggested in the ETSI TR 102 492-1 [2].

Table 2: Propagation parameters.

	Urban [6]	Suburban [6]	Rural [6]	ETSI [7]
d_0 (m)	64	128	256	15
γ_0	3.8	3.3	2.8	2.7
d_1 (m)	128	256	1024	---
γ_1	4.3	3.8	3.3	---

3.2 Vehicular to infrastructure radio channel model

In order to model V2I communications, we can use the WINNER (Wireless World Initiative New Radio) model [5]. The goal of WINNER is to develop a single ubiquitous radio access system adaptable to a comprehensive range of mobile communication scenarios from short range to wide area. WINNER II channel models set the basis for link level and system level simulations of local area, metropolitan area, and wide area wireless communication systems. The generic WINNER II channel model follows a geometry-based stochastic channel modeling approach, which allows the creation of an arbitrary double directional radio channel model. The channel models are antenna independent. The channel parameters are determined stochastically, based on statistical distributions extracted from channel measurement. Different scenarios are modeled by using the same approach, but different parameters.

WINNER II models can be applied to any wireless system operating in 2 – 6 GHz frequency range with up to 100 MHz bandwidth. The models support multi-antenna technologies, polarisation, multi-user, multi-cell, and multi-hop networks.

For vehicular networks we choose the B1 scenario from WINNER II, called "typical urban microcell". This scenario assumes a log-normal distribution of the shadow fading with a standard deviation equal to 3 dB. The path loss in LOS (Line of Sight) conditions is:

$$L_{\mathrm{LOS}}[\mathrm{dB}]=22.7\cdot\log_{10}(d[\mathrm{m}])+41+20\cdot\log_{10}\left(\frac{f_c[\mathrm{GHz}]}{5}\right),\quad 10\ \mathrm{m}<d<d'_{\mathrm{BP}} \tag{2}$$

$$L_{\mathrm{LOS}}[\mathrm{dB}]=40\cdot\log_{10}(\mathrm{d[m]})+9.45-17.30\cdot\log_{10}\left(h'_{\mathrm{BS}}\right)-17.3\cdot\log_{10}\left(h'_{\mathrm{MS}}\right)+2.7\cdot\log_{10}\left(\frac{f_c[\mathrm{GHz}]}{5}\right),\ d'_{\mathrm{BP}}<d<5\ \mathrm{km}$$

where d is the distance between the transmitter and the receiver in meters, f_c is the system frequency in GHz, and h'_{BS} and h'_{MS} are the effective antenna heights at the BS (Base Station) and the MS (Mobile Station), respectively.

The effective antenna heights h'_{BS} and h'_{MS} are computed as follows: $h'_{BS} = h_{BS} – 1.0$ m, $h'_{MS} = h_{MS} – 1.0$ m, where h_{BS} and h_{MS} are the actual antenna heights, and the effective environment height in urban environments is assumed to be equal to 1.0 m.

The breakpoint distance, d'_{BP}, is computed as follows:

$$d'_{\mathrm{BP}}=4\cdot h'_{\mathrm{BS}}\cdot h'_{\mathrm{MS}}\cdot\frac{f_c[\mathrm{Hz}]}{c} \tag{3}$$

where c = $3\cdot10^8$ m/s is the propagation velocity in free space.

The NLOS (Non Line of Sight) path loss model for scenario B1 is dependent on two distances, d_1 and d_2. These distances are defined with respect to a rectangular street grid, as illustrated in figure 3, where the MS is shown moving along a street perpendicular to the street on which the BS is located.

In figure 3 and in the next equation, d_1 is the distance from the BS to the centre of the perpendicular street, and d_2 is the distance of the MS along the perpendicular street, measured from the centre of the LOS street.

According to the previous situation, B1 scenario WINNER set the NLOS path loss as follows:

$$L_{\mathrm{NLOS}}[\mathrm{dB}]=\min\left(L(d_1,d_2),\ L(d_2,d_1)\right),$$

$$\text{where } L(d_k,d_l)[\mathrm{dB}]=L_{\mathrm{LOS}}(d_k)+20-12.5n_j+10n_j\cdot\log_{10}(d_l)+3\cdot\log_{10}\left(\frac{f_c}{5}\right),\ \textit{with } L_{\mathrm{LOS}} \text{ from (2)},\ k,l\in\{1,2\} \tag{4}$$

where n_j=max(2.8-0.0024d_k, 1.84), d_k and d_l in meters.

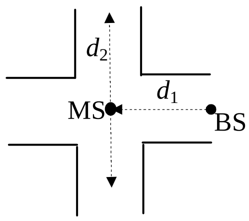


Figure 3: Geometry for d_1 and d_2 path loss model (from WINNER II model for B1 scenario).

4 APLICATIONS

Vehicular networks offer a wide range of applications. On a worldwide scale, numerous consortiums and projects have been set up and financed. In Europe, we would point out COMeSafety [6]. This Project counts on the members of the C2C CC consortium (Car to Car Communication Consortium), which brings together European automobile manufacturers. The sixth Framework Programme for investigation in the EU co-funds three projects in this field: SAFESPOT [7], CVIS (Cooperative Vehicle-Infrastructure Systems) [8] y COOPERS (CO-OPerative SystEms for Intelligent Road Safety) [9]. SAFESPOT is focused on the design and development of cooperative systems for road safety using V2V and V2I communications. CVIS project aims to design, develop and test the technologies needed to allow cars to communicate and network directly with the roadside infrastructure. Finally, COOPERS project focuses on the development of innovative telematics applications on the road infrastructure with the long term goal of a "Cooperative Traffic Management" between vehicle and infrastructure, to reduce the self opening gap of the development of telematics applications between car industry and infrastructure operators. The goal of the project is the enhancement of road safety by direct and up to date traffic information communication between infrastructure and vehicles.

5 CONCLUSIONS

This paper provides an overview of the state of the art in vehicular networks pointing out concepts and technologies in the vehicular communications field, and describing latest projects and research efforts.

The most important lines of investigation point to the evaluation of the performance and reliability of the systems in function of their index of penetration as well as advances in the study of radio channel communication and the adaptation of the Intelligent Driver Model (IDM) [10] to each and every situation and a more precise estimation of the given parameters, e.g. drivers' reaction times.

REFERENCES

[1] Dirección General de Tráfico, "La principales c*ifras de siniestralidad vial en España, año 2008"*, Observatorio Nacional de Seguridad Vial, 2009.

[2] ETSI TR 102 492-1, Part 1: Technical characteristics for pan European harmonized communications equipment operating in the 5 GHz frequency range and intended for critical road safety applications.

[3] ETSI TR 102 492-2, Part 2: Technical characteristics for pan European harmonized communications equipment operating in the 5 GHz frequency range intended for road safety and traffic management, and for non-safety related ITS applications.

[4] ECC Report 101: Compatibility Studies in the band 5855-5925 MHz between intelligent Transport Systems (ITS) and other systems.

[5] http://www.ist-winner.org/

[6] http://www.comesafety.org/

[7] http://www.safespot-eu.org/

[8] http://www.cvisproject.org

[9] http://www.coopers-ip.eu/

[10] A. Kesting, M. Treiber y D. Helbing, "Enhanced Intelligent Driver Model to Access the Impact of Driving Strategies on Traffic Capacity", *Traffic jams - dynamics and control of the Philosophical Transactions of the Royal Society A*, 2010.

Business Models of the new Digital Terrestrial Television in Spain

Víctor Correa Vázquez
victor.correa@coit.es

Abstract

The analogue switch-off completed in Spain during the month of April has terminated the process of migration to digital terrestrial television (DTT). The transition to the new technology has lasted more than five years and has completed with great success, generating in itself a business that according to official estimates has reached 12,000 million euros and has allowed the creation of 40,000 jobs.

Broadcasters will have to redefine their business models in a scenario with a larger number of factors (terrestrial, satellite or internet) in which the viewer is going to become scarce.

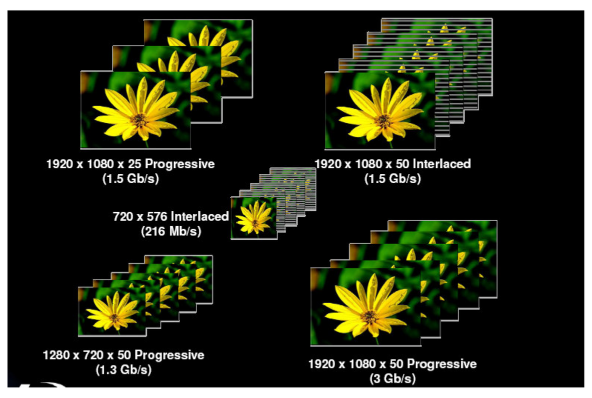
The fragmentation of audiences, the ability to access services and applications via DTT and paid content are presented as key elements to take into account new business models.

The first of these factors, audience segmentation, will enable advertisers to target a much more concrete way customers who are interested.

The second of them opens a market to explode new services and applications, offered directly by the "idiot box" to a greater potential audience than that of any other means.

Finally, pay TV market, begun by satellite and cable operators, in which Spain has a potential market that some estimates put the six million households may be attracted by sports channels (football mainly), high definition and 3D, to another sector of the audience.

High Definition



Introduction

This study intended to analyze changes in business models that may cause the features and functionality offered by new technology, namely fragmentation of audiences, interactive services and pay television.

Regarding the first factor, we must remember that the main source of income for the free TV is advertising, with revenues recorded under this heading in 2009, some sources estimated at more than 2,300 million euros in Spain.

With DTT, advertisers think they can address more specifically the target of customers who are interested, honing their goals. This selection of the customer is seen now in theme channels (sports, children, etc.) with targeted advertising to the target audience of the channel. Another strategy, aimed in this case to avoid fragmentation, is to broadcast simultaneous advertising with the issuance of the same ads in all programs of the same chain.

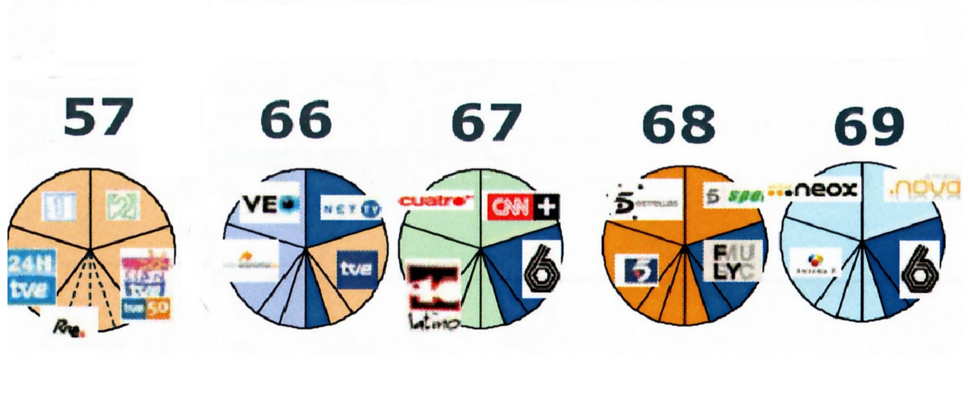
The second factor, interactive television, with frequent consultation services with little need for interactivity as the consultation of bank accounts, direct voting in contests, consultation of relevant information (weather, news briefs, sports scores, etc..), making representations to the request for a certificate, not to mention sports betting, has the advantage of reaching more customers without having to move them from their favorite place on TV.

Although there are government, especially local and regional level, which offer such services, the severe competition from the Internet, which also are turning the broadcast, is hampering the development of interactive TV, to the extent that data capacity of the multiple (1 Mbps per program) is being used for duties as channels of teleshopping, or other applications.

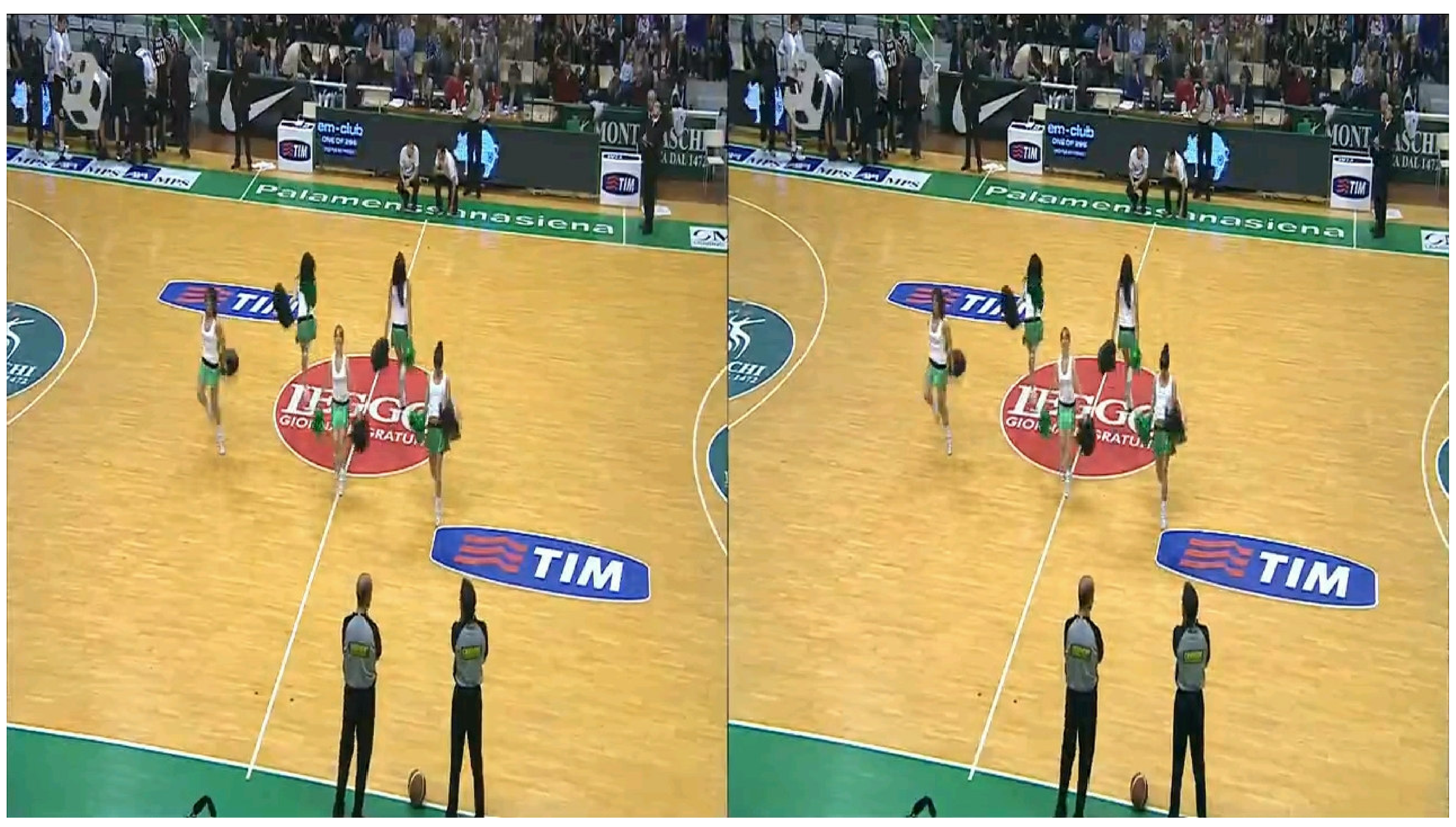
Finally, if the fragmentation of audiences is an established fact that has led to changes in business strategy, pay-TV with a good image quality (HD and 3D) and good contents is perhaps the factor with most influence in future business strategies.

Interactivity

Audience fragmentation. Theme channels



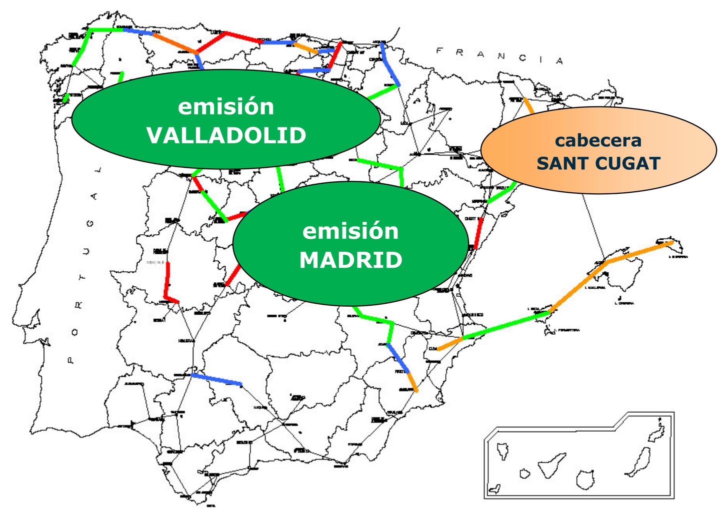
3DTV side by side



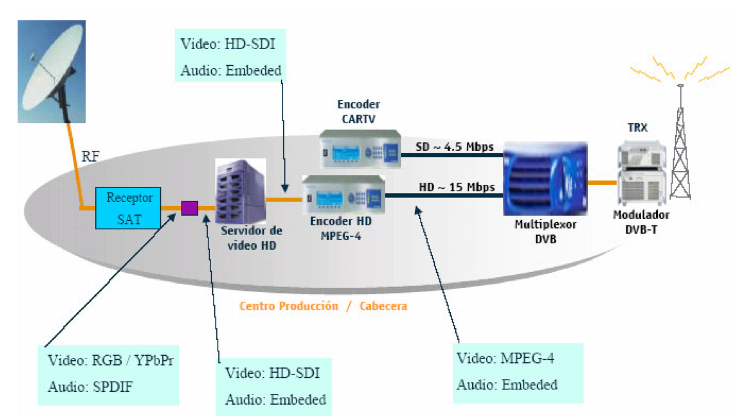
Experiences

The following are some experiences in high definition Tv and pay-tv done in Spain:

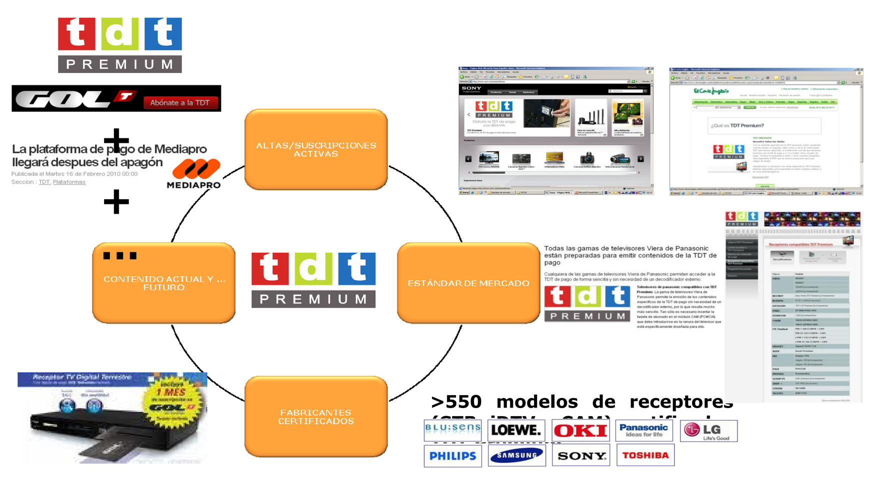
- Pilot HDTV Zaragoza, Mérida, Valencia and Murcia
- Pilot HDTV RTVE
- Premium DTT (digital pay TV)



RTVE Pilot



Mérida/Zaragoza HD Pilot



TDT Premium

Regarding the 3D television, the current inventory of the market situation is:

- BSkyB plans to launch the first 3D PPV channel in 2010

- Sony announces an agreement with FIFA to record and broadcast the World Cup in South Africa in 3-D film in over 7 countries

- TV3 is considering starting in 2010 to be tested in 3D.

- Abertis Telecom relays TDT 3D Asimelec Congress '10 in Seville with major TV manufacturers.

- Currently there is no 3D standard, which makes the economies of scale in TV sets and create a risk of fragmentation.

- Is expected to launch massive domestic TV 2012.

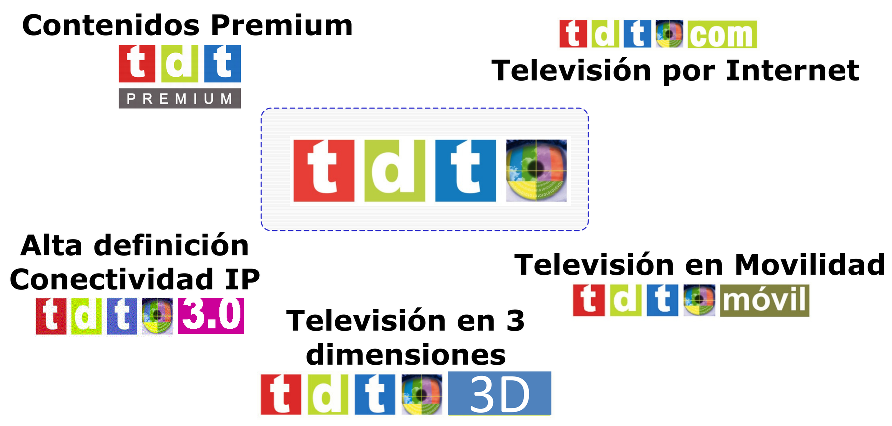
Future

Be planned three multiplex national, one regional and one for mobile TV under the Spanish National Technical Plan of DTT, using part of the spectrum released by the switch-off of analog TV broadcasts.

During 2010 will be deployed 3 new multiplex in a way that each private national TV will have 1 multiplex and TVE will have 2. The higher bandwidth will allow the development of HD and 3D services.

From "The content is King" to "The consumer is the Boss", he decides what to watch, when, how and where.

The DTT must remain competitive in this new scenario.



Conclusions

In summary, this analysis of the Business Models of the DTT in Spain leads to the following preliminary conclusions:

- **The audience fragmentation leads to a different model of advertising management with a better defined target (theme channels).**
- **Interactivity is pending to develop and will have difficulty in development because of the wide expanse of the internet connection.**
- **Pay-TV may have great development with a high penetration and low cost. The interest of the viewer depends on the content (football, movies, etc...) and the picture quality: HD, 3D, ...**
- **The new digital ecosystem leads to a convergence (technology and devices) unprecedented.**
- **The viewer becomes who really decide how, when and where he wants to see this content.**
- **Advanced services of DTT offer multiple possibilities for the audiovisual sector and society.**

A BRIEF HISTORY OF QOS

Edward Smith
BT, Bracknell ATE, Market St, Bracknell, Berks RG12 1JG,UK
edward.a.smith@bt.com

Keywords: QoS, DSCP, Shaper, Policer, Marker, Dropper, Traffic Engineering, Scheduling, TCP/IP

Abstract: The need for providing differentiated network performance for different classes of traffic is now well established, however this need has evolved over a number of years as remote access to computer applications has become more important and the economies of providing converged networks have become recognised. This paper will show how the provision of Quality of Service has evolved and how this enables the requirements of applications to be met.

1 INTRODUCTION

Differentiation between applications was not a significant issue for early communications implementations, since voice and data were provided over different networks, even if these were based on different TDM channels on the same physical infrastructure. In addition traffic data speeds were low by today's standards and traffic volumes were significantly lower.

This paper considers the evolution of differentiated service techniques, starting with the early poll response protocols used for data networks, taking the story as far as converged IP networks. The discussion will include IP QoS techniques, their implementation, traffic engineering and application requirements.

2 Early Implementations

Device prioritisation was an integral part of the C03 poll response protocol devised by ICL in the 1970s. This was initiated at link start up, with initialisation of the cluster controller by the host, followed by the assignment of categories to each station attached to the target cluster controller using a control packet sent from the host. This allowed the host to prioritise traffic from end devices by polling by device category. More sophisticated LAN based intelligent devices were able to support multiple applications across the serial link using the same mechanisms, thus allowing discrimination between interactive and batch traffic types.

Class of Service (CoS) (Randesi and Czubeki, 1992), a facility introduced by IBM in 1979, is provided at the Path Control layer of SNA sub-area networking and may be aligned to a virtual route across the network. A virtual route is a bidirectional logical path, which can be classified as high, medium or low priority. Differentiated services such as prioritisation, data sequencing and flow control on a virtual path may be superimposed on this.

Quality of Service was introduced in the 1984 recommendation for X.25 (Hura and Singhal, 2001), in the form of a throughput class (Thorpe and Ross, 1992), which is defined as the number of bits per second that can be transferred on a particular Virtual Circuit (VC). The default throughput class was derived from the speed of the access link and applies to all logical channels on the link. This is set up at subscription time is applied unless it is over-ridden during call set-up.

Tymnet (McDonnell-Douglas, 1988) implemented a connection orientated network, which uses a resilient supervisor to set up connections across the network, based on a set of throughput characteristics requested by the caller. Flow control was achieved using a technique known as gouging, which resembles the token bucket idea used in modern policers.

In Frame Relay (Atkins and Norris, 1995) the flow control mechanisms incorporated in X.25 are removed leaving the network open to congestion, which can lead to packet discard. Recovery is triggered when a network node detects congestion in one direction, traffic flowing along that directions will be tagged by setting the FECN (Forward Explicit Congestion Notification) bit in the frame header, while traffic flowing in the reverse direction is tagged by setting the BECN (Backward Explicit Congestion Notification) bit. The responsibility for recovery is delegated to the higher level protocols (usually TCP).

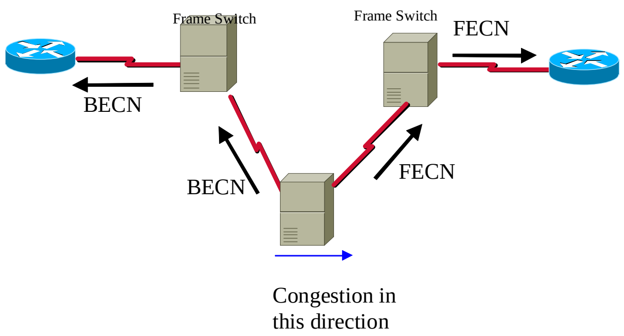


Figure 1 Frame Relay Congestion

The capacity of a Frame Relay PVC is normally defined by its Committed Information Rate CIR, which represents the throughput contracted for the service. Bursts smaller than a critical size known as the committed burst size are guaranteed transmission at the CIR, but if the burst size lies between this and a preconfigured excess burst size, then the traffic may be transmitted with the Discard Eligible (DE) bit being set. Any bursts larger than the excessive burst size are discarded on entry.

SMDS made use of a credit manager (Atkins, and Norris, 1995), where a connection is given a fixed "credit" to start with, which is usually enough to send the largest possible SMDS packet. As data is sent into the network, credit corresponding to the number of octets being sent is used up, reducing the overall credit balance. At the same time the sender continuously accrues credit at a rate determined by the access class. If insufficient credit is available then further packets will not be accepted by the network.

A number of different traffic classes are specified in the ATM standards, the most common of which is variable bit rate (VBR). Here QoS (Atkins and Norris, 1995) is described in terms of accuracy, dependability and speed. As a PVC is set up, the network determines that sufficient resources are available to support the required performance, without undermining the contracted levels of service for existing virtual circuits.

VBR traffic can be specified in terms of a Peak Cell Rate (PCR) and a Sustained Cell Rate (SCR) The former is the maximum rate, in cells per second, at which the user may send cells into the network whilst the latter is defined as the contracted cell rate that the connection can expect to achieve over a sustained period.

The dual leaky bucket algorithm, as required by the ATM Forum UNI specification, mimics a bucket with a hole in its bottom. The first bucket, shown below is dimensioned to allow traffic to enter the network, through the hole in its bottom, at the Sustained Cell Rate. If cells arrive faster than they can be serviced then data “overflows" into a second leaky bucket. Traffic flowing into the network, from this second bucket has its Cell Loss Priority bit set and will be discarded if congestion is encountered further downstream. Any overflow traffic from this bucket is discarded. Overflow continues until there is enough room again for new data to be accepted.

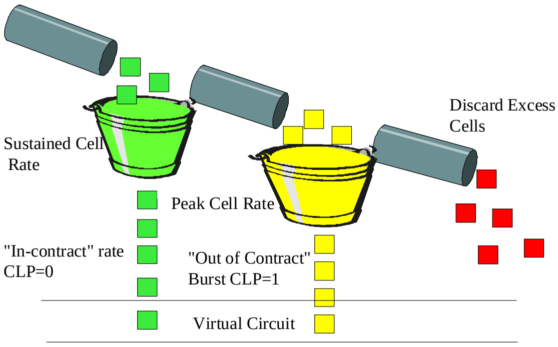


Figure 2: Dual Leaky Bucket Policer

Traffic shaping is a complimentary technique, which controls the metering of incoming traffic to reduce the occurrence of surges that could exceed the buffer capacity on any output port. The sending station delays the entry of large traffic bursts into the network, thereby ensuring a more constant flow of traffic and decreasing the likelihood of burst parameters being exceeded and packets being discarded, leading to a reduction in throughput.

3 QUALITY OF SERVICE IN IP NETWORKS

Early proprietary QoS mechanisms for IP networks tended to use three methods:

- Priority based queuing, that classified traffic into queues which were served in strict priority order. This could lead however to the higher priority queues starving the lower priority queues of bandwidth.
- Class based queuing splits the user traffic into a number of queues, with the available bandwidth on the link being served, being divided up between each queue on a pre-allocated basis.

- Weighted Fair Queuing, which using a time division multiplexing type mechanism, minimised the impact of intensive batch sessions, characterised by large packet sizes, on interactive traffic streams using fewer and smaller packets.

There are two significant architectures for achieving QoS,: the integrated services model or IntServ (Braden, Clark and Shenker, 1994) and the differentiated services model DiffServ (Nichols et al, 1998) . IntServ is the earlier of the two and requires the reservation of capacity across all network nodes to ensure that the right level of service can be provided. The Resource Reservation Protocol (RSVP) (Braden et al, 1997) is usually used to reserve sufficient resources to accommodate the desired load traffic. Traffic classification was performed on a per flow basis and this led to scalability issues, which inhibited wide adoption, however the underlying principles have been used in traffic engineering techniques for MPLS.

The DiffServ approach takes an approach based around the Per Hop Behaviour (PHB) of individual traffic types rather than individual flows, with traffic classified using settings in the TOS byte in IP Protocol header. The layout of this byte was originally defined in RFC 791 (Postel, 1981), refined in RFC 1349 (Almquist, 1992) and re-specified in RFC 2474 to use the 6 most significant bits of the TOS header to define the Differentiated Services Code Point (DSCP) to categorise traffic on a per hop basis (Nichols et al, 1998).

Blake et al (1998) describe an architecture for QoS enabled networks that has been widely adopted and which uses the following core components:

- Classification: The selection of packets based on the content of the packet headers according to defined DSCP field.
- Dropping: the process of discarding packets based on pre-set rules; to bring the stream into compliance with a traffic profile.
- Marking: the process of setting the DSCP code point in a packet header based on defined rules.
- Metering: the process of measuring for example the rate of a traffic stream selected by a classifier.
- Shaping: The process of delaying packets within a traffic stream to cause it to conform to some defined traffic profile.

This provides service differentiation in only one direction and is therefore asymmetric in nature.

Heinanen and Guerin describe a single rate three colour marker (srTCM) (Heinanen and Guerin, 1999a) and two rate three colour marker (trTCM) (Heinanen and Guerin, 1999b), that can be used as a component in a DiffServ traffic conditioner. Metering relies on information rates measured in bytes per second and burst sizes measured in bytes. In the single rate model traffic is marked using three measures: Committed Information Rate (CIR), Committed Burst Size (CBS) and Excess Burst Size (EBS). This is very similar to the mechanism described for Frame Relay. A two rate three colour marker uses two rates: Peak Information Rate (PIR) and Committed Information Rate (CIR), similar to the dual leaky bucket specified for VBR PVCs in ATM networks.

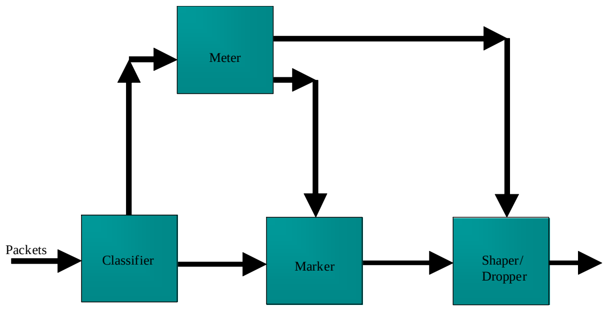


Figure 3: Logical view of QoS mechanisms from RFC 2475.

Jacobsen et al (1999) describe the per hop forwarding behaviour (PHB) for Expedited Forwarding (EF), to provide an end to end service giving low loss, low latency, low jitter assured bandwidth. This makes it necessary to ensure that the aggregate sees no, or very small, queues. This traffic type should receive its contracted data rate, irrespective of the intensity of the other traffic. The EF PHB allows unlimited pre-emption of other traffic, but its impact on other traffic needs to be controlled, for example using priority queue mediated by a token bucket limited policing mechanism.

There are three levels of drop precedence within each Assured Forwarding (AF) class, within the four AF classes (Heinanen et al, 1999). A congested Diff Serv node should try and protect packets with low drop precedence, by preferentially discarding packets with a higher drop precedence. This RFC gives an example where the drop precedence of a packet may be assigned using a leaky bucket policer, defined by a rate and a size, plus a committed and an excess burst size. Packets may be marked with low drop precedence if the number of tokens in the token bucket is greater than the excess burst size, medium drop precedence if the token bucket is not empty, but contains fewer tokens than the excess burst size and high drop precedence if the bucket is empty or the excess burst size is exceeded.

There are two principal ideas governing congestion management (Braden et al, 1998), these are queue management and scheduling algorithms. The former manages queue length through appropriate packet drop, whilst the latter determines which packet is transmitted next and is used primarily to manage the allocation of bandwidth across flows.

Tail drop, the traditional method of queue management, comes into play as soon as the buffer queue reaches a threshold value(s), normally the queue capacity. Packets often arrive in bursts and if a queue is near saturation multiple packets may be dropped, which in turn can result in lowered overall utilisation for a sustained period due to the TCP flow control mechanism causing the synchronised throttling back of flows.

"Active queue management" techniques, such as Random Early Detect (RED), drop packets before a queue becomes full, so that the end nodes can respond before buffers overflow. This allows routers to determine when and how many packets to drop by:

- Reducing the number of packets dropped (recovery from a burst of dropped TCP packets is harder than from a single lost packet).
- Providing a lower delay interactive service.
- Avoiding lockout behaviour.

The RED algorithm (Floyd and Jacobson, 1993) drops arriving packets probabilistically, with the likelihood of drop increasing as the queue grows. No packets are dropped until a pre-set minimum threshold is reached, whilst the number of packets dropped increases linearly from zero to the maximum drop probability between the minimum and maximum threshold. There after classic tail drop operates. RED controls the average queue size while still accommodating bursts of packets without loss. RED's use of randomness breaks up synchronised processes that lead to lock-out.

A refinement of this technique, known as weighted RED (WRED), allows preferential treatment to be given to traffic marked as (for example) AF21 over traffic marked AF22 and AF23, by applying a more aggressive discard policy (WRED profile) to the latter.

4 EXAMPLE IMPLEMENTATION

One of the key requirements of an MPLS network is the ability to support differential classes of service and permit the prioritisation and reliable delivery of different traffic types. Early implementations tended to use three classes of service, although more recently a six layer model based on DSCP markings has been deployed (Carter, 2005). Typical markings are:

- EF class, suitable for the transportation of packetised voice. This supports non-bursty, small data packets, with low cross network delay and minimal variation in cross network latency (this is known as jitter).

- Four AF classes, suitable for the transport of high priority data requiring low network delay and low probability of packet discard.
- DE (default) class, for lower priority traffic that needs less stringent delay and discard characteristics.

The standard model used by the BT's MPLS platforms offer the so-called hierarchical model shown below:

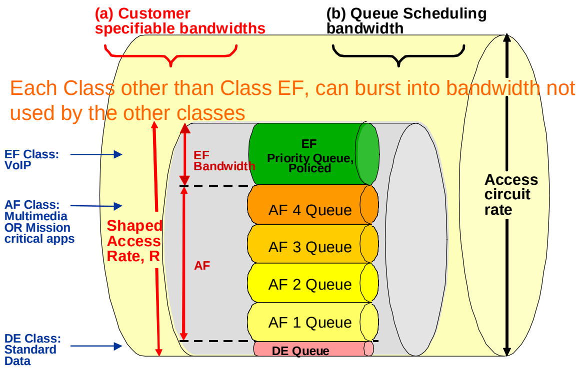


Figure 4: The hierarchical Class of Service model

This allows traffic to be prioritised effectively, allowing traffic from a lower class of service to make use of unused bandwidth in the higher classes of service. In this way the contracted CoS service levels can be delivered, whilst at the same time allowing optimal use of the available bandwidth.

5 TRAFFIC ENGINEERING IN IP NETWORKS

Traffic Engineering is the mechanism by which traffic crossing a meshed or partially meshed network is distributed to even out network loading. It aims to avoid creating points of congestion across the network, so that Service Level Agreements can be met with maximum reliability at minimum cost.

Layer 2 networks are engineered to route virtual circuits as efficiently as possible, allowing the traffic load to be more evenly spread across network trunks. In layer 3 based networks, such as MPLS (Awduche et al, 1999) traffic routing is governed by standard IP routing protocols, which direct traffic across the shortest available path irrespective of network loadings. This can lead to some trunk links being heavily loaded whilst others are under-utilised.

To overcome this, Traffic Engineering capabilities have been devised using the label switching capabilities defined by the IETF and documented in RFC 2702 (Rosen, Viswanathan and Callin, 2001). Spraggs (Spraggs, 2000) describes the design goals of MPLS traffic engineering as being to:

- Maximise the utilisation of network resources.
- Allow MPLS to replicate the traffic engineering functions of layer 2 networks such as Frame Relay and ATM.
- Integrate traffic engineering into layer 3.
- Route traffic flows based on the resources available in the network.
- Use either static routing or constraint based routing (CBR) to select the shortest path that meets the resource requirements of the traffic flow within constraints of available resources.
- Gracefully recovery from link or node failures that change the topology of the backbone by adapting to a new set of constraints.

The Traffic Engineering set up and maintenance mechanisms are commonly achieved by using RSVP as a signalling protocol, as shown in figure 5. Once set-up is complete, path selection is performed on the head end router, which determines the explicit path to be used for packet forwarding. Traffic is routed onto the tunnel, which is a unidirectional pipe, which only the head end router can put traffic onto.

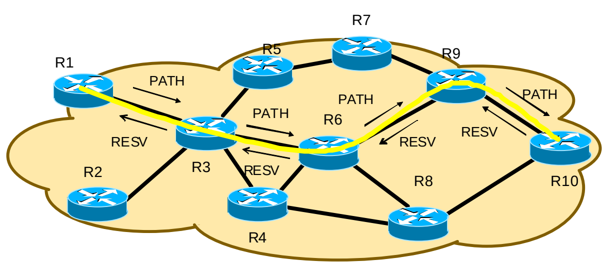


Figure 5: Tunnel Signalling, the route through the network is given by R1→R3→R6→R9→R10

6 APPLICATIONS REQUIREMENTS

In early IT developments, interactive and batch traffic were typically handled using different capabilities; with the former being usually allocated to terminals and cluster controllers and the latter to Remote Job Entry (RJE) stations. The wider adoption of IT across the business, the increased intelligence of end user devices and the advent of converged networks render this approach obsolete.

The responsiveness and availability measures for three traffic types: transaction orientated (response time based), throughput orientated (measured in kbps) and streaming orientated (specified by quality) are suggested in RFC 3729 (Waldbusser, 2004). Voice typically requires low latency and jitter and is therefore allocated to the priority queue in most QoS schemes, recommendations for handling the various implementations of video, which is normally allocated to its own AF queue are given in RFC 4564 (Bariarz, Chan and Baker, 2006). A second AF queue is typically allocated to carry signalling protocols leaving two AF queues available for high priority applications, with low priority applications being allocated to the default queue.

A model for developing services based around multiple applications types is given by McDonnell et al (2005), this shows the important of the service implementation in supporting such services. Such networks form the basis for initially managing and then guaranteeing the performance of critical applications (Dann, Gillam and Thornhill, 2005 and Wittgreffe and Khan, 2008). However, in general measurement of application performance is normally done as a trouble shooting exercise, rather than as regular monitoring. A comprehensive approach is costly and complex to achieve today, but may be possible with future developments (Smith, 2008). Investigation of methods for predicting and controlling network performance suggest that classical queuing methods under estimate bandwidth requirements because they don’t get burst sizes right. More sophisticated methods of evaluation exist, but these are felt to be too elaborate and ignore the fact that traffic volumes measured during network operation tend to be based on long term trends and miss bursts (Smith, 2006).

7 STANDARDS ACTIVITY

The IETF is continuing to develop QoS standards in the following principle areas:

- Clarification of existing standards.
- Developments in traffic engineering in areas such as: differentiated services across MPLS tunnelled environments, rapid re-routing around network failures and resolution of issues arising out of the scaling of link state routing protocols.
- The evolution of mechanisms to handle explicit congestion notification (ECN) to reduce the need for discarding packets.
- The communication of QoS parameters for remote access as a component of the Diameter Authentication, Authorisation and Accounting protocol.
- Specification of support for differentiated services, within MPLS label switching.
- The development of additional RMON MIBs for real time QoS reporting.

8 CONCLUSION

This paper has shown how the provision of Quality of Service has evolved and how this enables the requirements of applications to be met.

It has looked at the early poll response protocols used for corporate data networks and their migration towards more refined packet and cell based protocols. It then considered mechanisms available within Frame Relay, SMDS and ATM, showing how this has led to the development of queue control mechanisms such as policing and traffic shaping.

The evolution of these techniques in IP networks, moving from proprietary to standards based approaches was then examined, together with techniques such as traffic engineering. Finally the growing requirement for new applications and recent standards activity were described.

REFERENCES

Almquist, P., 1992. *Type of Service in the Internet Protocol Suite*, RFC 1349.

Atkins, J., Norris,M., 1995.*Total Area Networks*, John Wiley and Sons, Chichester.

Awduche, D., Malcolm, J., Agogbua, N., O'Dell, M., McManus., J.,1999. *Requirements for Traffic Engineering Over MPLS*, RFC 2702.

Bariarz, J., Chan,K., Baker, F. 2006. *Configuration Guidelines for DiffServ Service Classes*, RFC 4594.

Blake, S., Black, D., Carlson, M., Davies, E.,Wang, Z., Weiss, W., 1998 . *An Architecture For Differentiated Services*, RFC 2475.

Braden,R., Clark,D., Shenker,S., 1994. *Integrated Services in the Internet Architecture*, RFC 1633.

Braden, R., Zhang, L., Berson, S., Herzog, S. , Jamin, S. 1997. *Resource Reservation protocol (RSVP) – Version 1 Functional specification*, RFC 2205.

Braden, B., Clark, D., Crowcroft, J.. Davie, B., Deering,S., Estrin, D., Floyd, S., Jacobson, V., Minshall, G.. Partridge, C., Peterson, L., Ramakrishnan, K., Shenker, S., Wroclawski, J., Zhang,L.,1998. *Recommendations on Queue Management and Congestion Avoidance in the Internet*, RFC 2309.

Carter, S.F., 2005. *Quality of Service in BT's MPLS-VPN Platform*, BT Technology Journal, Vol 23, No 2, pp 61-72.

Dann, T., Gillam, J., Thornhill,D., 2005. *The Applications Assured Infrastructure*, BT Technology Journal, 24(4), 167-179.

Floyd, S., and Jacobson, V., 1993. *Random Early Detection gateways for Congestion Avoidance*. IEEE/ACM Transactions on Networking, V.1 N.4, p. 397-413

Heinanen, J., Baker, F., Weiis, W. , Wroclawski,J. 1999. *Assured Forwarding PHB Group*, RFC 2597.

Heinanen, J., Guerin, R., 1999a. *A Single Rate Three Colour Marker*, RFC 2697.

Heinanen, J., Guerin, R., 1999b. *A Two Rate Three Colour Marker*, RFC 2698.

Hura, G.S, Singhal, M., 2001.*Data and Computer Communications: Networking and Internetworking*, CRC, Boca Raton,

Jacobsen, V., Nichols, K.,Poduri, K., 1999. *An Expedited Forwarding PHB*, RFC 2598.

McDonnell Douglas, 1988, *Series 15 Networks: Tymnet Concepts*.

McDonnell, E.G., McLoughlin, M.J., Pretty, S.G., Slater, R,J .,Smith, E.A., 2005. *BT Transform – an IP-enabled business platform*, BT Technology Journal, 23(3), 121-136.

Nichols, K, Blake, S, Baker, F, Black, D., 1998. *Definition of the Differentiated Services Field (DS Field) in the IP V4 and IP V6 headers*, RFC 2474.

Postel, J., 1981. *Internet Protocol*, RFC 791.

Randesi, S.J , Czubeki, D.H., 1992. SNA IBM's Systems Network Architecture, Van Nostrand, New York,

Rosen, E.,Viswanathan, A.,Callin, R., 2001. *Multiprotocol Label Switching Architecture*, RFC 3031.

Smith, E.A., 2006. *Understanding performance issues in IP networks*, BT Technology Journal, 24(4), 167-179

Smith, E.A., 2008. *Engineering the network for applications and services*, BT Technology Journal, 26(1), 25-37.

Spraggs, S., 2000, *Traffic Engineering*, BT Technology Journal, 18(3), pp137-150.

Thorpe, N.M, Ross, D., 1992. *X.25 made easy*, Prentice Hall International (UK), Hemel Hempstead,.

Waldbusser, S., 2004. *Applications Performance MIB*, RFC 3729.

Wittgreffe, J., Khan,K., 2008. *Orchestrating end-to-end network and IT resources according to application level service level agreements*, BT Technology Journal, 26(1), 46-57.

A Model for QoS optimization at IMS Media Proxy Elements, in VoIP Next Generation Networks

Dimitris Kabilafkas, Ioannis Kordoulis, George Goulas and Konstantina Katsampani

Hellenic Telecommunications Organization (OTE S.A.), Research and Development Department,
Pelika and Spartis str., Marousi, Athens, Greece
dkabil@oteresearch.gr; ykordoulis@oteresearch.gr; gegoula@oteresearch.gr; katsampani@oteresearch.gr

Keywords: Quality of Service, Next Generation Networks, IP Multimedia Subsystem, Self-optimization, Self-configuration, Re-configurability, Traffic Management, Signaling Proxy, Media Proxy, Session Boarder Element.

Abstract: The present work introduces a model for Quality of Service (QoS) optimization on Application Layers, based on traffic measurements at media proxy network elements of the IP Multimedia Subsystem (IMS) Next Generation Networks. This technique is expected to coexist and interwork with the Layer 3 IP QoS methods, thus calling for integrated VoIP QoS management solutions. In the current paper, a functional application for either RTP packet delay calculations at VoIP calls that already have been established, or ICMP packet delay calculations, in case no VoIP calls are available, is described as well as the software design logic for VoIP session “parameter reconfiguration” at signaling and media proxy network elements, based on the measurement results of the initial application. Technically speaking, “parameter reconfiguration” in VoIP sessions means either the switching of the audio/video codec during an active call, based on SIP/SDP session renegotiation mechanism, or the initial selection of the appropriate audio/video codec, during the establishment process of a new VoIP call, based again on the same SIP/SDP session negotiation mechanism. In case that more than one network interfaces (for call routing) exist at IMS media proxy elements, an additional QoS technique resulted in Least Cost Routing is described, based on the same measurements on RTP or ICMP packets, for Best Effort interface switching of RTP traffic on either an already established or on a new VoIP call session.

1 INTRODUCTION

To meet the demands for converged packet infrastructures, Telecom providers are increasingly moving their networks toward an IP NGN-based architecture. There are two types of network deployment that currently enjoy the title of NGN: The first is the deployment of fibre into the local loop, either to the incumbent's street cabinet in conjunction with VDSL deployment or the deployment of fibre all the way to customer premises (typically apartment blocks rather than individual houses).
The second is the replacement of legacy transmission and switching equipment by IP technology in the core, or backbone network. This involves changing traditional telephony switches and installing routers and Voice over IP equipment to the network, end to end. The IP Multimedia Subsystem (IMS) is an architectural framework for delivering IP multimedia services to the customer. This framework architecture defines components, services and interfaces for next generation networks (NGN) and supports a wide range of services based on SIP protocol. As architecture it has the ability to deliver multimedia services that can be accessed from various devices through the IP or circuit switched networks.

2 IMS Reference Architecture Network Elements

IMS reference model is displayed in Figure 1. It defines the Network Elements as well as the communication protocols between the Network Elements of the proposed IMS reference architecture.

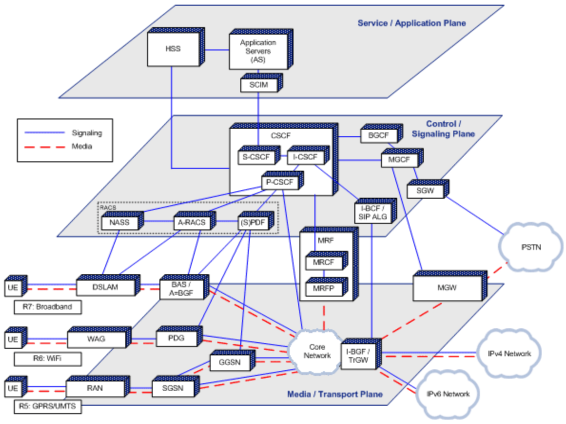


Figure 1: Schematic of IMS architecture reference model

The Call Session Control function (**CSCF**) is the heart of the IMS architecture and is used to process SIP signaling. Its main function is to provide session control for terminals and applications, using the IMS network. Session control involves the secure routing of the SIP messages, monitoring of the SIP sessions and communicating with the policy architecture to support media authorization. It is also responsible for interacting with the HSS. As it can be seen from Figure 1, CSCF can play three different roles: Serving-, Interrogating and Proxy-Call Session Control Function (S-, I- and P-CSCF).
HSS is the master database that contains user and subscriber information to support the network entities handling calls and sessions.
The **BGCF** (Breakout Gateway Control Function) selects the network in which PSTN breakout is to occur and within that network selects the MGCF.
Media Gateway Control Function (**MGCF**) is the central node of the PSTN gateway. The MGCF is responsible for controlling the media resources used when traffic needs to flow between networks using different media, typically between TDM network and an IP-based network.
IMS **MGW** (IMS Media Gateway) terminates bearer channels from CS networks and PS media streams and handles resources (echo cancellers, codes, etc.).
The **MRF** (Media Resource Function) provides the additional media resources complementary or fundamental for services. The MRF is split in to two functional parts: Multimedia Resource Function

Controller (MRFC) and Multimedia Resource Function Processor (MRFP).
SIP AS (**Application Server**) accepts requests and responses and is able to control, finish or initiate new SIP transactions. It interacts with other service platforms for the support of services, communicates with the HSS in order to obtain information about subscriptions and services and routes the session towards another user or network.

3 QoS Optimization model Architecture

System design consists of three separate modules, communicating each other through a common predefined space in Signaling and Media Proxy Network Element memory space (IMS SBC Memory Space), as displayed in Figure 2. The Index-Variable associated to this memory space in SBC is unique for every RTP stream of a SIP session, clearly named after the format printed below:

Index-Variable format name:
$(source.ip:port_destination.ip:port)

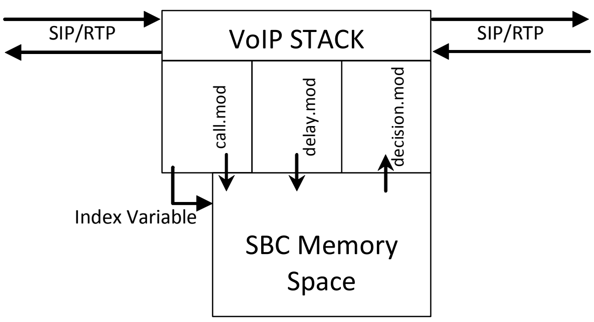
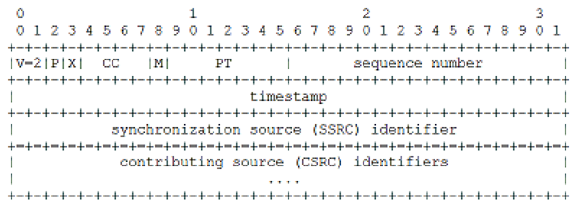


Figure 2: Modular System Architecture

Different RTP streams within the same SIP session at Signaling and Media Proxy Network Element are mainly separated from the source and destination port addresses, resulting in different Index-Variable memory names. The same goes for the RTP streams of 2, 3 or even more different SIP sessions between two end SIP devices. Obviously, RTP streams of different SIP sessions among different end SIP devices are separated from the different source or destination IP addresses, depending on the RTP direction, named after “source.ip” and “destination.ip” at Index-Variable format name.
In order to reach some realistic conclusions, avoiding complexity at the same time, the following assumptions concerning VoIP communication parameters have been made:

1. VAD is assumed to be disabled, in all VoIP communication streams.
2. Only two voice codec are being used in all VoIP sessions, G.711 and G.729.
3. It is assumed that there is not jitter in the way packets are being received at Signaling and Media Proxy Network Element.

The module that initiates the process for every SIP call session is named after **“call.mod”**. This module is being triggered at every SIP message containing SDP, gathering all the necessary information concerning the specific call session, such as source and destination IP addresses, source and destination port numbers, call-identification and finally the codec. Immediately after every SIP call establishment in IMS SBC Network Element (Signaling and Media Proxy Netowrk Element), “call.mod” module writes “call-identification” and “codec” values to system memory, at the specific space pointed by Index-Variable value of each SIP session call direction [named after Index-Variable $(source.ip:port_destination.ip:port) format].
“call.mod” is also being triggered at every SIP BYE message, clearing system memory space, pointed by Index-Variable of the SIP call session that is actually being terminated.



The delay module **“delay.mod”** is responsible for calculating the time delay between sequential RTP packets in a RTP stream of a call in progress, existing in IMS Signaling and Media Proxy Network Element (ie SBC).

Figure 3: Schematic of RTP

The specific number of RTP packets is defined by a Window Size variable, consisting of either the absolute number of RTP packets or the amount of time in msec in which the trace in SBC will take place, thus gathering a large amount of RTP packets, within this window time. From the RTP header, the only fields that are used are the “timestamp” and the “sequence number”.

The sequence number field, displayed in Figure 3, is used to determine that the received RTP packets are sequential, thus they can be used in calculating the reception delay. The timestamp field displayed in Figure 3 is used for correlating its own value with the delay value that is calculated in Signaling and Media Proxy Network Element, by subtracting the actual time value that sequential RTP packets, within the Window Size, are being received by the Media Proxy Network Element. The result is stored in the Index-Variable, have already been made by the previous module. The delay module continuously checks all the incoming and outgoing RTP traffic of the IMS SBC Network Element and stores the result in the appropriate Index-Variable.

The decision module **“decision.mod”** reads the output from the delay module (source.ip:port_destination.ip:port), where specific parameters such as call id, codec and packet delay have been identified for each call. The purpose of this module is to initiate (or not) a new SIP/SDP Negotiation Session, as displayed in Figure 4, using different codec order, according to the configurable threshold value that is predefined for the specific type of application, meaning that different packet delay values will be used for audio, video etc. In case that QoS mechanism is being triggered by the decision module, the SBC will issue a re-Invite SIP message to the other party in order to renegotiate the codec characteristics of the session, setting codec G.729 first one in the SDP Codec Order Field.

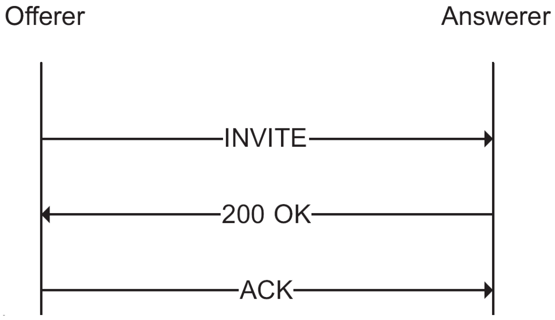


Figure 4: Schematic of SIP/SDP Session (re)Negotiation

If B party accepts the new codec (presented first one in codec order priority), the call will normally continue, using the new one codec, otherwise it has two ways to evolve: either the call will continue, using the old one codec (G.711) or the call will be terminated. “Decision.mod” is being triggered when a new Index-Variable is being issued by the call.mod: “source.ip:port_destination.ip:port”.

4 Traffic Management based on LCR

In case that SBC holds more than one network interfaces for the same destination networks, mainly used either as load balancing for traffic or redundant links for failure states, an additional QoS method can be performed, based on the triggers of the previous modules, resulting in using the alternative network interface for routing VoIP session RTP packets. In this Least Cost Routing scenario, SBC reroutes RTP traffic of a specific VoIP session to the alternative interface, when delay module calculates a delay value more than the predefined threshold value. This mechanism runs only after VoIP session parameter reconfiguration, described in previous paragraph, has already being executed once, meaning that voice codec for the specific VoIP session has already being changed from G.711 to G.729, according to the mechanism described earlier.

5 Conclusions

This paper introduced a new QoS technique that aimed at “parameter reconfiguration” in VoIP sessions at IMS SBC Signaling and Media Proxy Network Elements, meaning either the switching of the audio/video codec during an active call, based on SIP/SDP session (re) negotiation mechanism, or the initial selection of the appropriate audio/video codec, during the establishment process of a new VoIP call, based again on the same SIP/SDP session negotiation mechanism.
Active calls are not being terminated during codec switching process; however an audible gap may exists during the transition time, as a result of lost or delayed packets due to the adaptation process of the active call to the new session parameters (new audio/video codec).

An additional QoS technique resulted in Least Cost Routing has also been described, based on the same measurements on RTP packets, for Best Effort interface switching of RTP traffic, on either an already established or on a new VoIP call session.

REFERENCES

IETF RFC 3261——SIP: Session Initiation Protocol, (Authors: Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., Schooler, E.) Internet Engineering Task Force (IETF), 2002.

IETF RFC 2327——SDP: Session Description Protocol, (Authors: Handley, M., Jacobson, V.) Internet Engineering Task Force (IETF), 1998.

IETF RFC 3264——An Offer/Answer Model with the Session Description Protocol (SDP), (Authors: Rosenberg, J., Schulzrinne, H.) Internet Engineering Task Force (IETF), 2002.

A. Raake, Speech Quality of VoIP – Assessment and Prediction. UKChichester, West Sussex: Wiley, 2006.

R. Ramjee, J. Kurose, D. Towsley, and H. Schulzrinne, “Adaptive playout mechanisms for packetized audio applications in wide-area networks,”in Proc. 13th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM ’94), USA–Toronto, CA, 1994, pp. 680–688.

ECTA comments on NGN public policy

http://en.wikipedia.org/wiki/IP_Multimedia_Subsystem

DISRUPTIVE MECHANISMS IN THE QOS PROVISION

The way for a new decade of Quality in Communications

Josu Bilbao[†], Igor Armendariz[†], Pedro Crespo[‡]

[†]*IKERLAN-IK4 Technological Research Centre,Mondragon, Spain*

jbilbao@ikerlan.es,iarmendariz@ikerlan.es

[‡]*CEIT-IK4 and Tecnun (University of Navarra),Donostia-SanSebastián, Spain*

pcrespo@ceit.es

Keywords: Quality of Service (QoS), Network Coding, Linear Network Coding, Random Network Coding, Routing, Throughput, Delay, Network Monitoring.

Abstract: With the XXIst century, a new mathematical flow emerges from the Information Theory field with the aim of reaching new up-bound limits in the use of data networks resources, in order to maximize the transmitted information. Mentioned research field has been named as *Network Coding*, and provides innovative mechanisms for the provision of Quality of Service (QoS) in Next Generation Networks. Network Coding is taking shape as a key agent in the transformation of the communications towards a higher efficiency in the network resources use, and reaching a higher QoS from a prism of different application fields. The treatment of flows based on transmitted information criterion instead of the transmitted bits, provides a new conceptual frame, intended for encouraging new communication ways in the near future. Present work describes the main mathematical underlying concepts, which will comprise an emerging research field that will determine the future communication networks design. Present paper aims to highlight the advantages of using Network Coding instead of classic Routing strategies commonly used in packet switching networks. Provided benefits are justified by analyzing the opportunities and challenges born in a new framework characterized by considering communications based on the transmitted information instead of transmitted flow.

1 INTRODUCTION

During the last decade, communications requirements of data networks have experimented an important growth. With the "*Future Internet*" topic, as the European Commission has named the evolution of data networks, and the bloom of networked nodes in the "*Internet of Things*", the need to share communication networks resources with higher requirements has experimented an exponential growth.

Scientific research community has made a huge effort in order to define innovative mechanisms with the aim of managing the shared network resources to achieve the QoS (*throughput, delay, jitter, bit-error-rate, etc.)* response needed for the new application and services.

Traditionally, packet switching networks have used the different information data flows as separated elements, even if they had to share common network resources. However, with the XXIst century emerged a new Information Theory field known as *Network Coding (NC),* which breaks the assumptions of separating information flows.

First, basic concepts related with NC will be described from a conceptual and mathematical point of view. Subsequently, NC codifying methods will be depicted. The following paragraph benefits and cost of NC will be highlighted. Afterwards main incipient application fields will be described, as well as the roadmap for future application fields that are being analyzed by the research community.

Finally, conclusions of present work are highlighted.

2 New paradigm for qos provision

This new Information Theory field is based on very simple mathematical concepts, but constitutes a new generation in the treatment of dada networks with a revolutionary improvement in the response of packet switching networks and its QoS. Network Coding allows nodes to accomplish linear combinations among packets, rather than the restricted functions of replication and forwarding that are typically allowed in conventional store-and-forward mechanisms.

Network Coding theory has been studied and boosted from different theorical fields: Algebraic (*Koetter and Médard*), combinatorial (*Fragouli and Soljanin*), information theoretic (*Alshwede and Yeung*) or the linear programming frameworks have tried to tackle the problem and establish the fundamentals that will help in the development of new practical way to understand the communications, and reach new achievable limits.

2.1 Routing versus Combining

During decades, communications have been based on point-to-point solutions and the data transmission paths were solved with routing algorithms. However, Network Coding modifies the communications concept in new generation networks in favor of network resource efficiency.

In point-to-point communications, store-and-forward strategy could not be the optimal one when the following actions are required:

- Multicast transmission.
- There are 2 or more information independent to be transmitted both in unicast or multicast.

Maximizing multicast rate using routing is NP-hard, while Network Coding design is in Polynomial time. It must also be taken into account that NC could be applied in any system that could be modeled as a network.

The following figure will help in the basic explanation of the concept behind NC operations. Considering the scenario depicted in Fig 1, node A pretends to transmit two information units (m_1, m_2) in multicast to F and G sink nodes.

With routing strategies (left image), it can be seen that D node becomes a *bottleneck*, given that it can resend both incoming packets in the same time. Hence, routing requires two transmissions to multicast (m_1, m_2) to F and G sinks. However, in the right image, using Network Coding, a packet combination ($m_1 \oplus m_2$) is done with incoming symbols, transmitting the multicast information to both F and G sinks in the same time. As both nodes receive also m_1 and m_2, they can decode the source information.

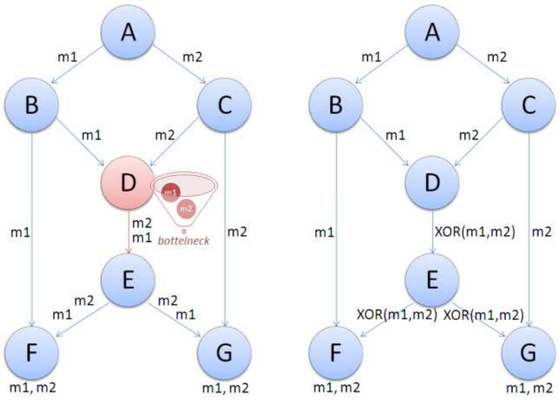


Figure 1: a) classical routing (left) and b) Network Coding (right).

The achievable limits of unicast transmissions have been defined more than 50 years ago, while multicast achievable limits are already an open study field.

2.2 Max-Flow Min-Cut Theorem

The basic theorem that describes the achievable ratio in a communication network is the so called "*Max-flow Min-cut Theorem*", and is the concept that opens the door to the irruption of Network Coding theory. This theorem defines the limit of information that can be transmitted in multicast over a given network. It can be resumed that:
Given a communication network with point-to-point links, we can define a directed graph $G = (V, E)$, where V represents the nodes set, and E represents the edges set in G (the point-to-point links). Lets asume that $|E| < \infty$, so that, $|V| < \infty$.
We define s as the source node and t as the sink.
The *Min-cut* between the s node and t node can be seen as the bottleneck between both nodes.
Therefore, a *cut* between the source and the receiver is the set of edges whose removal disconnects source from receiver (Yeung, 2008). Hence, min-cut represents the minimum value. It is evident that the max-flow between s and t will not exceed the capacity of mentioned min-cut.

Theorem: Being G a graph with a source node s , a sink node t and a rate constraint R ; then the *max-flow* value from node s to node t is equal to the capacity of the *min-clut* between both nodes.

This theorem means that the *min*-cut capacity is always achievable.
If we define ω as the ratio of information to be multicast from source s to sinks $t_1, t_2, \ldots, t_L$ in a given network G with rate constraints R , intuitively we can prove that:
$\omega \leq maxflow(t_l)$

And if we take into account several receivers l =1,2,3,…L,
$\omega \leq min_l \; maxflow(tl)$

The result indicates that information can be send at a maximum rate of min-cut, and defines the limit for a multicast transmission.
It can be highlighted that *max-flow* is always achievable based on routing mechanisms when there is a lonely sink. However, when there are 2 or more sink nodes, we can achieve the *max-flow* using Network Coding thanks to the theorem explained in the following paragraph.

2.3 Network Coding Theorem

Lets consider a multicast scenario over the graph $G = (V, E)$ where there are h source flows over the same node S $(S_1, S_2, \ldots S_3)$, transmitting simultaneously multicast information to N receivers $(R_1, R_2, \ldots R_N)$.
We assume that G has unit capacity links and that the *min-cut* value between source and each sink is h.

Theorem: Being the graph $G = (V, E)$ with unit capacity edges, h parallel sources with multicast information placed in the same vertex, and N receivers. Then, there exist a transmission scheme over a large enough finite field $\mathbb{F}_q$, over which, intermediate nodes combine the incoming symbols over $\mathbb{F}_q$, and delivers the information of source nodes to each receiver at a rate equal to h.
From the *min-cut max-flow theorem*, we know that there exist h disjoint paths and each R_j receiver

will receive information from h sources routed over the h disjoint paths. However, when multiple receivers use the network simultaneously, their paths may be overlapped. Hence, they must share the available network resources, and it will reduce the communication ratio. The NC Theorem indicates that, if intermediate nodes are allowed to combine incoming information flows, then each sink receives the information at the same ratio as if it could have an exclusive access to the network resources.

Over this theorem it can be added that it is sufficient for the intermediate nodes to accomplish linear operations (add and multiply) over a finite field $\mathbb{F}_q$, what is known as *Linear Network Coding*.

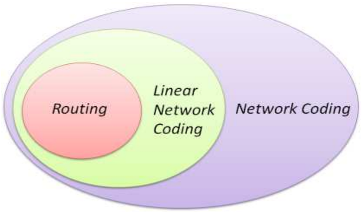


Figure 2: Scheme of different communication strategies.

The use of linear combinations over intermediate nodes can be interpreted as a particular case (subset) of using Network Coding. In order to reduce the complexity, linear combinations over a finite field $\mathbb{F}_q$ are preferred faced with non-linear, though the alphabet size could be significantly reduced as we move from linear to nonlinear solutions.

With Linear Network Coding, intermediate nodes, multiply the incoming symbols with coefficients in $\mathbb{F}_q$ and are linearly combined with the rest of the incoming symbols.

2.3.1 Achievable rate

Following the described theorems and the graph of Figure 1, it can be appreciated that the use of NC provides a bigger achievable rate region. Lets assume that messages m_1 and m_2 are transmitted at rates r_1 and r_2 respectively. From previously defined theorems it can be highlighted that the achievable region of both strategies is as shown in the following figure.

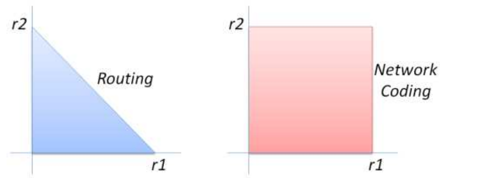


Figure 3: Achievable rates without (left) and with NC.

3 Coding methods

3.1 Linear Network Coding

Network Coding, instead of a simply store-and-forward, combines incoming flows in intermediate nodes delivering one or several outgoing flows. Concretely, the use of Linear Network Coding is very similar to the one used in figure 1, substituting the XOR $(\oplus)$ by any linear combination of incoming symbols in a finite field $\mathbb{F}_q$. This gives more flexibility in the combining options.

The information received by sink nodes do not depend on the specific received packets, but in the reception of a enough number of linearly independent packets. This way, a sink node could complete the information received even if a given packet don't arrive because of a lossy link or a specific node has been disconnected.

3.1.1 Coding Process

Lets consider an acyclic graph $G = (V, E)$, with unit capacity edges, a source node $s \in V$, and a set of receivers $T \subseteq$ V. Taking into account the *max-flow min-cut* theorem, we consider the broadcast capacity to be h.

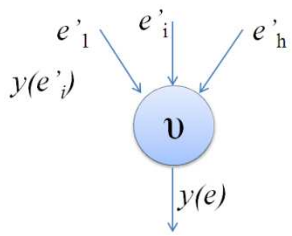


Figure 4: NC combining incoming information flows.

Being e' the incoming edges in node v, input simbols are identified as $y(e')$. Hence, the resulting outgoing symbol is a linear combination of the incoming symbos:

$$y(e) = \sum_{e':Out(e')=v} m_e(e') \cdot y(e')$$

If we take into account the source symbols that $y(e'_i) = x_i$ for $i = 1, \dots, h$.

Therefore, the outgoing symbols are: $\sum_{i=1}^{h} g_i(e) \cdot x_i$

The vector of dimension h formed by the coefficients $g_i(e)$, is called the *global encoding vector*.

3.1.2 Decoding process

The sink node t that recieves the simbols from h sources, will receive $y(e_i)$) symbols so that:

$$\begin{bmatrix} y(e_1) \\ y(e_2) \\ \dots \\ y(e_h) \end{bmatrix} = \begin{pmatrix} g_1(e_1) & \cdots & g_h(e_1) \\ g_1(e_2) & \cdots & g_h(e_2) \\ \vdots & \ddots & \vdots \\ g_1(e_h) & \cdots & g_h(e_h) \end{pmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_h \end{bmatrix}$$

Hence, in reception it can be defined a linear equation system as $Y = G_t \cdot X$, where G_t constitutes the transition matrix formed by the used coefficientes.

In order to decode the original information (x_i) sink node must solve the equation system with the received information. This equation system will be solvable if the matrix compound by the coefficients used in the coding process (and received by the sink) is *Full Rank*. Said in other way, $\det|G_t| \neq 0$. Hence, the G_t matrix will be invertible and equation system will be solvable:

$Y = G \cdot X$

$X = G^{-1} \cdot Y$

One of the key points is that in order to decode the received information, it is no needed the reception of every packet. The lost of a given packet can be supplied by the reception of any h packets that allow the matrix to be invertible in reception, and therefore solve the equation system in sink nodes. This advantaje of NC allows important advances in *Fault-Tolerant* systems.

3.2 *Random Network Coding*

One of the most critical points in Network Coding is the selection of coefficients used to complete linear combinations among incoming flows, taking into account the need to complete a equation system with the requirements described in previous paragraph.

One of the simplest alternatives is to prefix the coefficients before. However, these kind of solutions are less practical in dynamic scenarios where nodes could leave or connect temporarily. This becomes the selection of coefficients a complex problem.

As a practical approach, (Ho and Médard, 2007) propose an alternative based on a random selection of coefficients following the next theorem:

Theorem: Considering a graph $G = (V, E)$, being V the set of nodes, E the set of edges and N receivers, where coefficients used in coding are selected uniformly at random from a field $\mathbb{F}_q$, a multicast transmission can be accomplished if a sufficiently large field ($\mathbb{F}_q$) is used.

In (Fragouli and Soljanin, 2007), it is proved that if we chose a Finite Field $\mathbb{F}_q$ with $q > N$, the probability that N receivers could decode the h sources is at least $\left(1-\frac{N}{q}\right)^{\eta}$, where $\eta \leq |E|$, and being η the maximum number of coding nodes used for each sink.

The advantage of this algorithm is that it does not need neither centralized information nor used topology knowledge, because random coefficients selection is done independently in each coding node. Hence, this method provides a practical and affordable computational cost coding algorithm.

4 Benefits of nc

Due to space limitations in the present paper, it is not possible to give a deep survey about the benefits provided by the use of Network Coding. However, it worths highlighting some of the key points related with the advantages that NC offers.

The use of NC was initially focused on the throughput and shared resources use optimization for multicast transmissions. Nevertheless, research community has started to find new potentialities in order to deal with security, delay reduction, network monitoring, topology inference, multicast flow distribution, disk array systems, fault tolerant networking solutions, reliable communications development, P2P networks, network on chip solutions, electronic circuit design, etc.
Present paper will describe some of the aforementioned advantages with the help of a few application fields description.

5 Cost of using NC

When the viability of applying NC is analyzed, as there is no advantage for free, it must be taken into account the cost, and needed resources to perform the linear network coding for multicasting. Two kind of cost can be highlighted:

- *Cost related with design*: the cost of selecting the coefficients for the linear combination, find the coding opportunities and manage changes in dynamic environments.
- *Operational Complexity*: Cost associated with the coding and decoding process. The coding process of h information flows, requires $\theta(h^2)$ operations over a finite field $\mathbb{F}_q$. The decoding process needs to solve $h \times h$ lineal equations, so that it needs $\theta(h^3)$ operations with a Gaussian elimination algorithm.

Hence, the use complexity of using NC is conditioned by the finite field size over it operates, and the number of coding nodes.

6 Application Fields

6.1 Content distribution (P2P)

One of the first application fields where the use of NC algorithms is proposed is the content distribution over Peer-to-Peer networks. A P2P network is based on distributing a given content over different nodes in the network, so that these nodes could recover the content working in a cooperative and decentralized way.

The source of content, divides it into n blocks, so that the initial content is divided into $m_1, m_2, \dots m_n$. However, instead of distributing original fragments, coded versions of these fragments are delivered to different nodes placed in the network, and these nodes will send linear combinations to connected nodes.
One of the most key points is that using random NC, topology knowledge is not required.
Blocks $m_1, m_2, \dots m_n$ are represented as symbols over a finite field $\mathbb{F}$, and coding process will define linear combinations of these blocks as described in:

$$y = g_1 m_1 + g_2 m_2 + \cdots + g_n m_n$$

The coeefficients $g_1, g_2, \dots, g_n$ have been selected in a random way over $\mathbb{F}$ (*Random Network Coding*).
The main advantage of applying Network Coding in P2P networks is based on the reception process. A P2P network could be modeled as a graph $G = (V, E)$, being V the set of nodes and E the set of edges. For a correct decoding, sink node needs to receive n linearly independent coded blocks, so that the inverse of the matrix is achievable and therefore, the linear equation system is solvable. This fact allows a higher reliability of the system faced with the lost of a given packet, and constitutes a flexible scheme in face of node disconnection.
Another interesting point is the analysis of the time required by the receiver to dispose of the content in a P2P network. Being $s \in V$ the content source, and $t \in V$ the receiver, we have that, when the base field is large enough, with probability close to 1 (Yeung, 2008), the theorem is accomplished:

$$maxflow(t_i) \geq n$$

Hence, taking into account the *max-flow min*-cut theorem, the time required by the receiver node t will be minimum, with a probability close to 1.
Thanks to the use of NC in P2P networks, maximum bandwidth efficiency is achieved, as well as the maximum reliability and flexibility. However,

a higher computational complexity is required compared with store-and-forward strategies. For the block linear combinations, $\theta(h^2)$ operations in $\mathbb{F}_q$ are needed, while decoding based on matrix inversion and Gaussian elimination requires $\theta(h^3)$ operations, being h the number of input flows.
As a practical implementation of this strategy, we can refer to MSCD (*Microsoft Secure Content Distribution*), also known as *Avalanche* (Gkantsidis, 2006).

6.2 Wireless Networks

Wireless networks are mainly characterized by the shared physical media and a very dynamic topology. These two characteristics are the base for the chance to apply the benefits of NC. In fact, wireless sensor networks are shaped as one of the first practical applications using NC mechanisms, as NC helps to exploit the broadcast nature of wireless.

6.2.1 Opportunistic Coding

The NC mechanisms require the existence of coding opportunities that will exploit the benefits described on mentioned theorems. One of the most illustrative examples of the opportunistic coding is the use of NC in wireless networks. The following figure shows that a store-and-forward mechanism needs 4 steps to communicate symbols from 2 nodes.

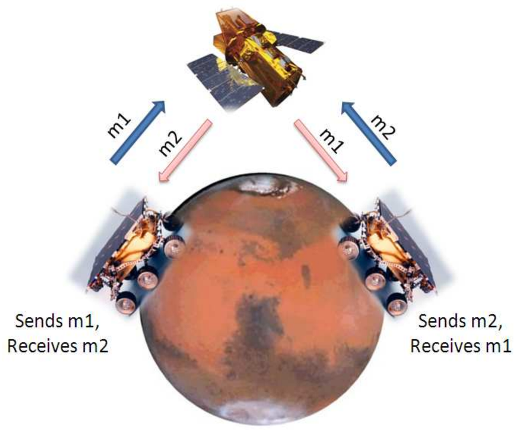


Figure 5: Wireless communication with store-and-forward.

However, using NC mechanisms, source nodes need only 3 steps to send the same information than the previous example (m_1 and m_2), thanks to the opportunistic routing concept. Combination of packets in the intermediate node generates an outgoing symbol that, thanks to the broadcast nature of wireless, sends the whole information to both nodes. Once they receive the coded symbol, they can extract the original information as they know the symbol they have sent and can therefore decode the complementary information.
It can be highlighted that with a very simple coding mechanism, a 50% resource save could be achieved in the downlink connection.

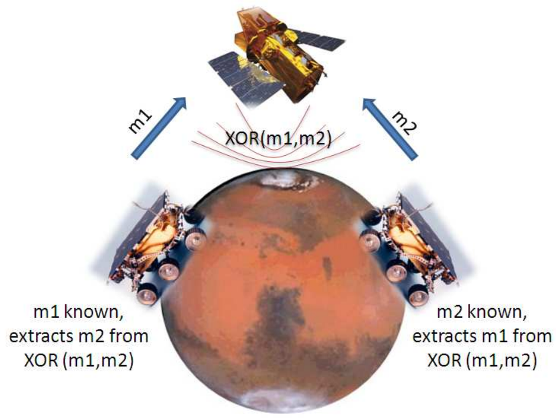


Figure 6: Wireless communication with Network Coding.

6.2.2 NC on wireless network throughput gain

The process of quantifying QoS benefits related with the use of NC in wireless network is not trivial, because benefits depend on the existence of coding opportunities. However, several research work have been done with the aim of analyzing the number of transmissions required to transmit a given information among the different nodes connected by a wireless network. In (Fragouli and Soljanin, 2007), it is described a relation between the minimum number of required transmissions to send an information unit to every node in the networks, as a function of the number of transmission needed, with and without Network Coding, T_{nc} and T_w respectively.

Being n the number of nodes in the network, it can be proved that:

$$\frac{T_{nc}}{T_w} = \theta\left(\frac{1}{\log n}\right)$$

It can also be concluded that:

- Network Coding, under topology *no-knowledge* assumption achieves the same performance as routing with a *"perfect knowledge"* of the topology.
- Under the hypothesis that neither of the schemes know the topology, Network Coding reduces the number of transmissions by a factor of $\log[(n)]$.

6.2.3 Energy efficiency in wireless sensor networks

From the previous remarks related with the different number of transmissions needed to transmit a given information to the rest of the nodes, it can be observed that the use of NC implies and implicit energy consumption reduction. This fact has an increasing interest in the *Internet of Things* and the new challenges for the *Smart Building* and energy efficiency (*Green IT*).
Hence, one of the main application fields of NC will be the wireless sensor networks, as the energy consumption reduction will lead to smaller batteries, or reach longer autonomy with same batteries.

6.2.4 Adaptability in dynamic networks

Thanks to the use of Random Network Coding, the knowledge of the network topology is not needed, which is a very important advantage in dynamic environments, as is the case of wireless networks. A packet loss due to the disconnection of a given node will not be critical, as the reception of enough number of packets from the rest of the nodes, containing linear combinations of different information flows (with random coefficients), will be enough.

6.3 Security with NC

In this paragraph we will consider 2 different scenarios.

6.3.1 NC against Eavesdropping

Thanks to the philosophy of combining incoming flows in order to generate outgoing flow, NC can provide an interesting shield against anyone who can "hear" one of the links (in case of routing, listening to a single link could be enough to know the whole information).
Considering a graph where the *min-cut* of a given receiver is h , we have that $s = (s_1, s_2, \dots, s_k)$ is the information associated with the k symbols wanted to be transmitted securely, and $y = (y_1, y_2, \dots, y_h)$ the random variables associated to the coded symbols.
Lets consider $z = (z_1, z_2, \dots, z_\mu)$ to be the random variables corresponding to the flows intercepted by a non desired listener.
In the described scenario, a theoretically secure scheme is considered when s is completely decodable by y , and is not conditioned by the information in z that the intruder will have listened.
We need to accomplish the following entropy conditions:

$$H(s|y) = 0$$

$$H(s|z) = H(s)$$

In order to achieve required security, NC is based on the condition that the intruder should have to intercept all the links, and therefore all the information, and not only a given single link, as is the case when routing mechanisms are used.

6.3.2 NC against Byzantine attack

Given the nature of combining information in the intermediate nodes, NC mechanisms are especially vulnerable to byzantine attacks. The attack over a single node could compromise the whole information transmitted to the rest of the nodes, which will propagate the problem combining again the altered information with correct one. The way to counteract byzantine attack will be introducing redundant information that will allow "*error correction coding*" mechanisms.
Another alternative, described by (Fragouli and soljanin) is based on the use of a *Vandermonde* parity check matrix, H , and distribute this matrix to every receiver using a secret secure channel, for which the required bandwidth would be very small compared to the one required by the main transmission.

7 Future Applications

Even that the first NC applications have been focused on the multicast network resource optimization, several new application fields emerge from the research community taking advantage from the mechanisms provided by NC. Subsequently, some of the application fields on study are described briefly.

7.1 Network Monitoring, tomography and link loss inference

Network Coding can be used to monitor a given network precisely with reduced overhead, analyze its topology, measure the packet loss ratio, delay and link errors, among others. This fact is inherited from the concept of combining information flows in the intermediate nodes. These intermediate nodes will allow the receiver identify the trace of the paths followed by the probe packets.

In (Yeung, 2008), it can be appreciated the use of NC to achieve a precise estimation, valid for any topology and having used minimum resources. In (Fragouli and Soljanin, 2007) NC based tomography inferencing algorithms are described. Allowing linear combinations over a finite field in the intermediate nodes, the network tomography inference is achievable.

7.2 Switching

A switch is a device used in order to connect different network segments at OSI link layer. As a basic limitation of switches, a given input can't transmit different information to different outputs in the same time.

The use of Network Coding can optimize the operating speed of a switch, allowing opportunistic coding to combine incoming information in the input queues and accommodate multicast traffic patrons. The following figure shows and example of using intra-session NC in a theoretic switch.

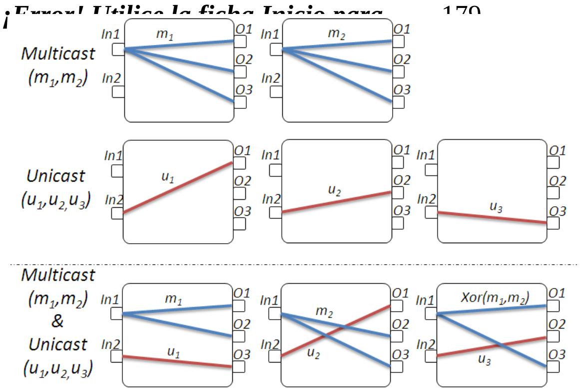


Figure 7: Multicast and unicast transmissions over a traditional switch (above) and a Network Coding switch (below).

In the previous figure, the upper configuration is the classical one, which requires 2 time units to transmit multicast content and 3 time units for the unicast flows, which makes a total of 5 time units required. However, in the lower part of the figure, a NC switch configuration can be seen. The use of *intra-session* NC allows to send the same information of the classical alternative using only 3 time units.

It can be highlighted that the use of NC in switches eliminates the need for speedup or to accelerate the intern performance of the switch, which will be traduced in a cheaper hardware.

7.3 Network on Chip

In the literature it can be found first steps of a new research line that starts proposing the use of NC techniques in the VLSI circuit design. The goal is the design of faster electronic, given that NC provides mechanism to increase the intra-chip link efficiency and the optimization of communications between logical units.

Advantages could also be extended to the PCB routing schemes with conventional electronic, achieving cheaper circuitry.

7.4 Physical layer NC

The concept of Network Coding can be applied over different OSI layers. One of the incipient application field with more promising future is the one called "*physical layer NC*". Relay nodes don't have to decode the information in each node, but they only transmit to the receivers enough

information for their decoding process with the aim of extracting the original information.
Signals from different nodes become into interferences that can contribute to the transmitted information, instead of being considered as destructive interference. The physical layer NC has a close relation with cooperative communications research line and the space-time coding algorithms.

7.5 SVC multimedia streaming

One of the applications with the support of the research community is the transmission of multimedia content using the benefits of NC. In particular, one of the most interesting concepts is the use of hierarchical mechanisms.
The goal is to prioritize packets that allow the reception of basic information to rebuild the multimedia content in sink nodes. Hence, even if several packets are lost, the reception of a sufficient number of packets will assure the arrival of required set of lineal combination of the original information, in order to rebuild the content with a higher reliability than using classical mechanisms.
In (Lui and Xiao, 2009) the applicability of NC in SVC (*Scalable Video Coding*) video systems is foreseen. In this case, if the total amount of linear combinations (or packets) are received, highest quality content can be rendered at receivers. Mentioned research work uses a *hierarchical NC*, that will allow that different nodes receiving the same multicast information could rebuild the content with the quality for which they have access in their profiles or depending their subscription.

8 Conclusions

During the last decade, QoS requirements for different services have suffered an exponential growth. With the start of the new millennium, a new theoretical research flow emerges, where data flows are combined in the intermediate nodes in order to accommodate the requirements of resources for different traffic patrons, mainly multicast, over existing networking infrastructures.
The born of a new theoretical base is encouraging the scientific community to research and re-invent communication concepts that we thought they have reached their capacity bounds and optimize the QoS response.
The need to optimize the use of the shared resources is a problem repeated periodically in the history. In the 80’s, the cost of memory was dramatically expensive. Hence, research community was focused on the development of memory management systems. During those days, it was easy to predict that the cost of the memory would decrease drastically with the years, so that memory management wouldn’t be required any more in future PCs. However, history has taught us that, even the cost of memories have been decreased enormously, the memory management is still required, as well as the complex hierarchy relations among them, because the memory requirements have suffered a growth an order of magnitude bigger.
Therefore, and even that in future we will have communication networks with higher bandwidths, for sure we will need to optimize even more the resources, because requirements will also increase at least in the same order. We should not forget that the stone era did not finish because humanity was run out of stones, but because we found new tools more effective.
From the QoS in future networks point of view, Network Coding mechanisms will be the foundations that will shape the change of the way we understand the communications in the future, focused in the end-to-end information transmission instead of data flows corresponding to this information.
The born of a wide variety of application fields justify the huge potential of Network Coding, which is already know as the “*Modern Theory of Communications*”. Theory that could change the way we understand QoS in the mid-term future. However, a lot of research work is still needed to define the impact Network coding will have in real applications.

ACKNOWLEDGEMENTS

This work was supported in part by the Strategic Research Project program of the Government of the Basque Country (Spain), through the project TEReTRANS, Expt: IE8-221.

9 References

Ho, T., Médard, M., Shi, J., Effros, M. and Karger, D. 2003. "*On randomized network coding*". In 41st Annual Allerton Conference on Communication Control and Computing.

Yeung, RW. 2008. "*Information Theory and Network Coding*". Springer.

Fragouli, C., Soljanin, E., 2007. "*Network Coding Fundamentals*". Now Publishers, Boston-Delft.

Fragouli, C., Soljanin, E., 2007. "*Network Coding Applications*". Now Publishers. Boston-Delft.

Gkantsidis, C.,2006. "*Network Coding for Large Scale content distribution*". Microsoft Research.

Lu, J., Xiao, S., and Wu, C. 2009. "*Efficient protection scheme for SVC content based on Network Coding*". 5th International Conference on Information Assurance and Security. IAS, Xian. Pp. 585-588.

RUBENS: DELIVERY OF ON-DEMAND CONTENT WITH IMPROVED CUSTOMER EXPERIENCE

Experimental and Techno-Economic Validation

Rob Smets[1], Mark Prins[1], Pieter Nooren[1], Steven van den Berghe[2], Koen de Schepper[2], Rafael Huysegems[2], Thomas Monath[3], Colin Pons[4], Gerardo Garcia de Blas[5]

[1]Dutch Organization for Applied Scientific Research (TNO), Brassersplein 2, 2612CT Delft, The Netherland, [2]Alcatel-Lucent Bell Labs, Copernicuslaan 50, 2018 Antwerp, Belgium. [3]Deutsche Telekom Laboratories, Goslarer Ufer 35, 10589 Berlin, Germany. [4]Royal KPN, Regulusweg 1, 2516AC Den Haag, The Netherlands. [5]Telefónica I+D, Emilio Vargas 6, 28043 Madrid, Spain. Email: rob.smets@tno.nl .

Keywords: QoE, QoS, Control Plane, Transport, Techno-Economics, Target Costing, MACTOR

Abstract: The position of operators in the end-to-end content delivery value-chain is challenged by flattening revenues and increasing cost for installing more bandwidth. Technical enablers implemented in a demonstrator allow content to be delivered with QoE using less peak-bandwidth. This new way of content delivery provides new retail and wholesale opportunities as users are willing to pay for QoE and proper technology exists.

1 INTRODUCTION

The delivery of content from content creators to consumers, requires the involvement of aggregators, distributors and manufacturers of customer premises equipment (CPEs). In the value chain they construct, distributors a.k.a. network or service operators are faced with flattening revenues and an increasing cost of ownership, while being for a large part responsible for the quality of experience of the end-user who expect SD to HD quality, choice in (popular)content, responsiveness of the service and a user-friendly electronic program guide (EPG).
The introduction of VDSL2, FttX and Docsis3 technologies are likely to shift traditional bandwidth bottlenecks in the access networks towards the core and service platforms where they traditionally can only be mitigated by over provisioning of bandwidth. The RUBENS [1] concept discussed in Section 2 aims at solving this potential bottleneck with better efficiency, at the same time adding value to content delivery. Section 3 discusses the experimental validation of the RUBENS concept, describing the architecture and technological enablers. The techno-economic validation of Section 4 shows the feasibility and how QoE can be 'sold'. Conclusions can be found in Section 5.

2 The RUBENS approach

2.1 Leaving Classic Content Delivery

IPTV and VoD front services deliver content linearly. During playback, using multicast or real time streaming, sustainable bandwidth needs to be available where small fluctuations in bandwidth can be tolerated by appropriate buffering. Media delivery over the Internet relies on more extensive buffering using technologies like progressive download. Switching between content results in waiting time to build up a large enough buffer and loss of bandwidth because existing buffers of arbitrary size are flushed without its content being viewed. Caching chunks of over http retrieved content does improve content delivery but cannot meet hard-guarantees when it comes to QoE due to cache misfiring and lack of signalling. The RUBENS approach however includes signalling, the user's QoE perception and is not hampered by cache misfiring. [14]

2.2 Properties of Content

RUBENS benefits from properties of content: Content has a degree of *popularity*, can be *personalized* to a end-user profile, can be delivered with a certain *lead-time* and in different *qualities*, and content can be *segmented*. These properties can be used to steer both the QoE perceived by the end-user, and confine the resources needed for content delivery to a set of end-users. In the RUBENS approach, content is divided and transported in semantic segments. These segments can range from a few seconds up to several minutes in duration. They represent reusable fragments such as a camera-shot, a scene, a news topic, a learning item, etc. Because these segments are independent of each other, they can be reused by intermediate nodes for different personalized play-outs.

2.3 QoE and Usability

The QoE perception of end-users depends on the objective QoS parameters used to deliver content together with the subjective perception of the end-user. RUBENS uses popularity, personalization, lead-time, and quality as QoE influencing parameters to control the level of congestion the delivery of media experiences in the network. This way a sufficient QoS is met and meeting the user's subjective perception of the service is what remains. A usability study [15] has been conducted by exposing a panel of users to a mock-up implementation of the service. The members of the panel were confronted with content delivered using different combinations of QoE influencing parameters. Afterwards they were interviewed regarding their experience. From the interviews some basic rules could be extracted:

For paying (premium) users:

- zero lead-time and high quality mandatory
- wide choice and personalized content

For non-paying (standard) users:

- exchange lead-time for high-quality
- smaller choice for better quality and lead time

2.4 Segmentation and Playmaps

To maximise the possibility for content-sharing within the Rubens network, these reusable semantic segments will be transported from server to access node, spanning the entire access and metronetwork. Just before the content is served to the consumer, (e.g. in the access node) the semantic segments are glued together to form again viewable content. The assembly is not a simple reconstruction of a static, unified content-item. Instead, it is a personalized and interactive remix of available semantic segments, tuned to the QoE preferences and profile of the individual consumer. The interactive remix can be tailored according to his interests, age, knowledge on a specific topic, previously seen content, etc. Even the duration of the content can be tailored to meet the available time of the consumer.

In order to steer the reassembly in a way that is meaningful for the end-user, metadata is made available per content-item to form a segment building plan: the playmap.

3 Implementation

3.1 Architecture

RUBENS uses a cross-layer approach to link different functions across OSI layer and between data and control plane. Figure 1 shows the architecture of the implementation.

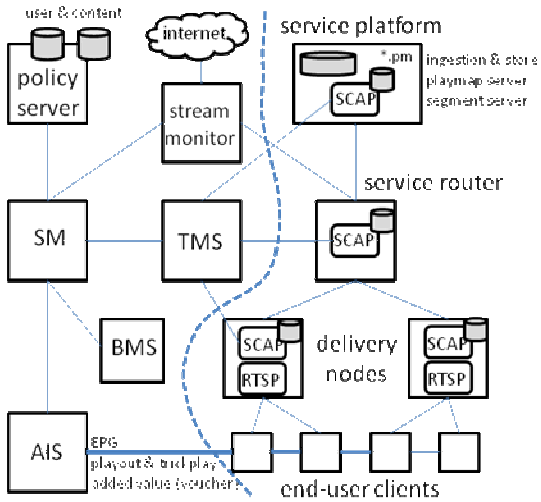


Figure 1: Architecture of RUBENS implementation

The data plane consists of a tree-like architecture connecting end-user's clients via delivery nodes and a service router to the service platform and the Internet. In this context deliver nodes represent access multiplexers (AMs) and the service router represents an IP transport network.

The control plane provides an application interface (AIS) to the clients and via transport management subsystems (TMS) and stream monitor to the data plane. The data plane is responsible for delivering the content to the end user.

3.2 Data Plane

The data plane consists of a centrally placed service platform containing a segment server, and a playmap server. Delivery nodes close to the end-user contain an RTSP streamer and facilities to handle RTSP requests of end-users that contain a reference to a playmap and retrieve this playmap from the playmap server. The delivery node will interpret the playmap using the QoE parameters passed on in the RTSP request. To start streaming the delivery node places request for appropriate segments towards the segment server using a novel transport protocol called SCAP (Shared Content Addressing Protocol). The SCAP protocol packets contain a L4 globally unique content identifier (GUCID) for the content requested and transported. It consists of the L3 host identification (IP address) combined with a 128 bit server-unique L4 content byte address. Bytes are individually addressable, but requested in a consecutive range with start and size parameters of the content that should be delivered. A stream-deadline (identified by a start deadline for the first byte, and an end deadline for the last byte in the range) is added in a request packet, which will allow the server to identify the deadline for each response byte and thus also for each packet. The GUCIDs of the bytes in the range allow intermediate SCAP nodes to detect duplicate content ranges. The deadlines will turn the disadvantage of requesting content upfront (additional load on the network because of more sharing of bandwidth, even if the delivery is not urgent), into a scheduling freedom advantage. The scheduling freedom allows earliest deadline first scheduling of the content, giving priority to urgent content, and resulting in fair deadline sharing instead of fair bandwidth sharing (as in TCP). [16]
The first SCAP node is also located in the delivery node and the first segment that is requested has a start deadline of 0 second (content needs to be played immediately) followed by subsequent requests where the start deadline increases, matching the end deadline of the previous segment, as these subsequent segments are needed at a later time.
When a SCAP node receives a request for a content range it will first look into its buffer and the parts of the requested content that are not present will be requested towards the upstream SCAP server. If a content range is available on a SCAP node it will be serviced from this node. Eventually requests reach the segment SCAP server where service is always possible.
During the traversal of requests SCAP nodes maintain information on deadlines to meet, workload to perform and service rates provisioned and communicate this to the control plane (TMS).

3.3 Control and Management

The control plane is centred around service management (SM). The main task of SM is to handle incoming requests from AIS. These requests originate from end-users requesting Electronic Program Guides, the play-out of content (incl. trick play features) and added value services like the use of vouchers that allow zero-QoS internet content to be played out using the QoE enabled RUBENS network.
SM interacts with a policy server to learn on QoE rules that may be user or content specific to set boundaries for the QoE space. This applies both to EPG and play-out requests. Next, TMS is accessed to learn on the state of the SCAP transport network and may restricts QoE space in which the content is played out.

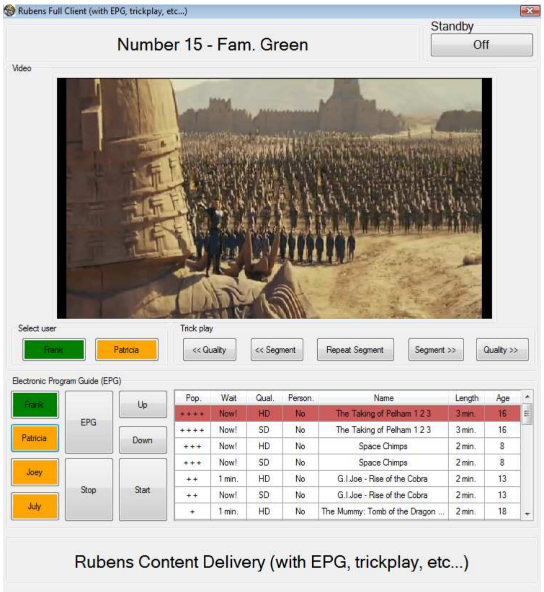


Figure 2: Demo client with EPG and play-out screen.

Bandwidth Management Subsystem (BMS) is consulted for last-mile bandwidth resources in support of RTSP. Last-mile fibre connections do not further restrict QoE space, copper-based technologies like ADSL may further impose restrictions on the QoE parameters used in the delivery of the content. When SM has finished computing information for the EPG, or for the play-out of a specific requests, it forwards all QoE parameters to AIS which assembles the EPG XML file or the URL for the play-out of content and returns this to the client. The client (Figure 2) can now display an updated EPG to the end-user or access the content with proper QoE parameters to the delivery node. Alternatively, when users activate a voucher, the stream monitor is instructed to filter internet traffic for specific content request signatures of e.g. YouTube in the user's internet traffic. When a match is found SM will try to find a replica of YouTube's content in the RUBENS database and when available the control plane returns an appropriate URL to the end-user.

3.4 Admission

RUBENS QoE admission control starts with the request for EPG information when choice in content may be temporarily restricted to popular content for non-premium users due to network congestion detected by TMS. The more congestion is imminent the stronger the limitation. Premium users do not experience this limitation as they are full-paying customers. Additionally the EPG reflects long term and course estimates on network performance in the communication of available quality, personalisation and lead-time. When content is selected for play-out, a more accurate calculation is made based on the most recent available data in TMS. In contrast to traditional admission behaviour (accept or reject) RUBENS will always present the end-user with a viable option.

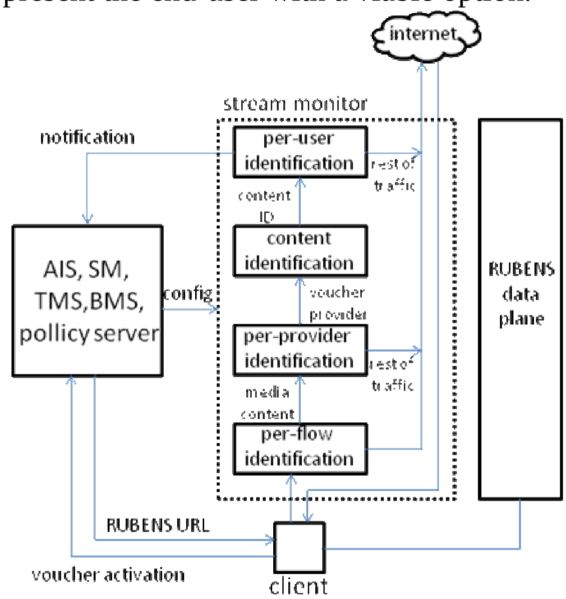


Figure 3: User initiated network-based admission.

In addition to measurement based admission control RUBENS supports network based admission control (Figure 3) to provide an alternative for low-end internet media delivery. Users may activate a “YouTube” or “Joost” voucher (e.g. provided to them as part of a sales offer) which enables network based access. The YouTube or Joost “EPG” can now be used to select content and when the request is made over the internet, the stream monitor will detect the request and instruct SM to deliver the media item using the RUBENS content repository (using original content and possibly personalized advertisements related to the sale offer) and QoE enabled SCAP delivery platform.

4 Techno-Economics

This chapter gives an overview of the techno-economical evaluations regarding the QoE Market and functionalities which were analysed.

4.1 QoE Related Economic Challenges

Changing user behaviour from television (broadcast) to video-on-demand (unicast) [1] will result in rising network congestion during peak hours and less bandwidth available to individual users. In addition to costly over-dimensioning of the network [2] to avoid a decrease in perceived quality by the user the RUBENS QoE steering technology can be used to smoothen the increase of network capacity and thus expenses. To validate this, customers must be willing to pay (1) for Internet services and content but also (2) for higher levels of quality. Both are needed to drive QoE technologies but willingness to pay for content does not imply willingness to pay for quality.

4.2 QoE Market Scenario Analysis

In order to determine if there is a market for QoE, a scenario analysis [6], [7], [8] was used to check for drivers, showstoppers and to assess the QoE environment. The results were derived from an analysis conducted in two workshops with 13 participants and one questionnaire with 20 respondents of which the majority are experts. Five main scenarios were identified (likelihood rated independently for each scenario in brackets):

- *QoE Heaven* (50%): The network capacity meets its limits but there is a friendly regulatory environment and a high demand and high willingness to pay for services. There is a general alignment and cooperation within a value network.
- *Immediate action required* (47%): The network capacity operates at its limits. The demand and willingness to pay for reliable services is high but QoE technology is not available. There exists a high need for investment in network capacity.
- *Dead Zone* (24%): Politics and regulators promote increase of network capacity resulting in network capacity to exceed the demand largely. Customers turn away from online services.
- *New Offer* (53%): QoE technology is ready but customers have limited appetite for QoE functionality, hence a low willingness to pay for those functionalities.
- *Regulation crashes the party* (42%): Politics and regulators promote the increase of network capacity but all other conditions favor the introduction of QoE.

Based on the scenarios some recommendations were developed:

Joint efforts between value chain members are promising to open up the pre-mature QoE-enabled content market potential. When QoE technologies are introduced in the network they open up retail and wholesale market opportunities that should be leveraged instantly. Consumers must be “sensibilized” and re-activated to appreciate high quality content delivery services strongly directed to fully exploit retail market potential. Vertical integration should be considered as a possibility.

4.3 Retail Market Potential

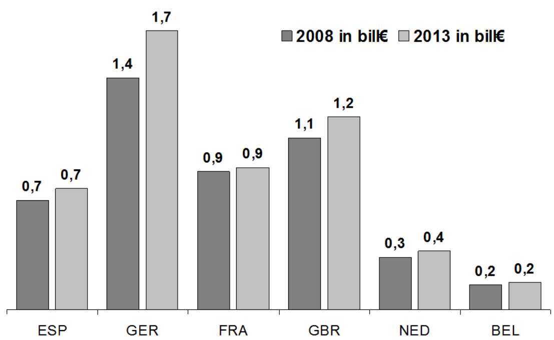


Figure 4: Estimation of overall annual potential market gains in analysed EU countries if QoE functions are introduced

A QoE market potential analysis for the retail market has been performed based on different sources: studies and reports, e.g. [9], [10], as well as consortium partners' data. The study was conducted based on forecasts for the QoE household penetration, data about Internet and TV consumption and demographic data taking into account the expected potential market shrinkage in case QoE functions are not introduced and the expected potential market revenue gains. The detailed model description can be found in [5]. In addition to the estimation overall annual potential market gains when QoE functions are introduced (Figure 4), a major finding is that an implicit demand for QoE does exist of 421M€ in 2013.

4.4 Competitive Environment of QoE Market

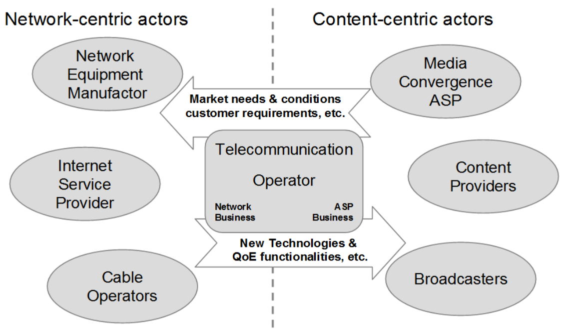


Figure 5: Telecommunication Operators (telco's) may act as mediating role in the QoE

The MACTOR (Matrix of Alliances and Conflicts: Tactics, Objectives and Recommendations) method was introduced to analyze and identify the relevant actors and key strategic issues (Figure 5) [11].
The results of the analysis are exhaustive and are described in detail in [4]. The role of telco's as both network centric actors and content-centric actors bears chances for the market as well as for telco's. They might act as promoter, facilitator and mediator on issues that separate the network- and content-centric groups. Through the telco's the content-centric group can communicate customers' needs, intentions and requirements to the other actors and the Network-centric group can demonstrate and promote QoE technologies to the content-centric actors. Telco's benefit from a two-sided market: a end-user related retail market (B2C) and a wholesale market (B2B).

4.5 Target Costing

The Target Costing analysis is applied to the IPTV service built on the RUBENS concept and technologies (Figure 6) and assumes that the component which serves customer needs the most should get most resources [5].
The customer needs were obtained from secondary sources such as studies, reports (e.g. [12]) and expert estimations to fill in missing customer requirements and their estimated costs and rate the individual needs in terms of importance for customers and thus the allocation of resources (Figure 7).

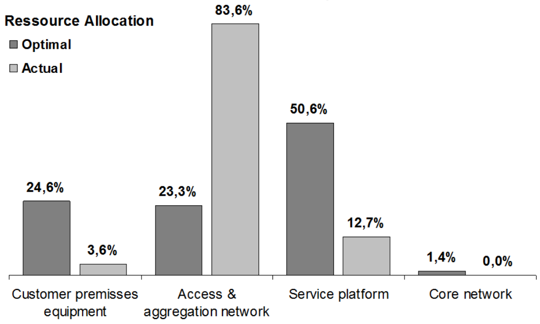


Figure 6: Optimal and actual allocation of resources for IPTV product components based on customer requirements and actual costs calculations.

The prices for a single IPTV offering were calculated (25% of a triple play package, Figure 7). For the German market the prices were verified to vary between 8 and 17€ [12], 13€ was taken for further calculations using the MUSE reference model [3] for network resource estimation and an existing network penetration curve [13] for 50.000 subscribers. Based on these assumptions, total revenues (10% discounted) of about 1.90 M€ in the seven year period result.

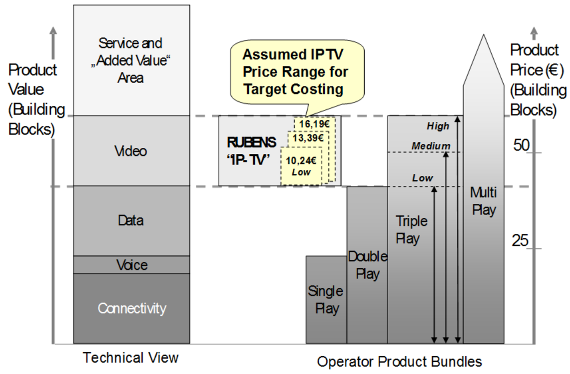


Figure 7: Assumed RUBENS IPTV product and related price range in context to the product value and operators product bundles.

Figure 6 shows, a dramatic shift in resource allocation from the access & aggregation (RGW, DSLAM, BRAS, etc.) to the service platform and CPE. The degree of intelligence to be introduced into aggregation network components and transport product functions has to be analyzed carefully.

4 ConClusions

RUBENS has demonstrated the feasibility of QoE enabled media delivery through implementation of a demonstrator and techno-economic evaluation. Technical enablers like segmentation, construction of playmaps and a novel transport and access control concept allow manipulation of QoE parameters under different QoS regimes. Telco's are provided with an improved position in the value chain when traditional bandwidth investments are replaced with investments in value adding QoE delivery platforms.

REFERENCES

[1] http://wiki-rubens.celtic-initiative.org/index.php/Main_Page

[2] J. Bryant and M. B. Oliver, "Media Effects - Advances in Theory and Research", 3rd Edition ed. New York: 2008.

[3] IST MUSE. (2008, 08.02.2010). DA3.3 - Economical guidelines and decisions in BB access (results of use cases). http://www.ist-muse.org

[4] T. Heger, *et.al*., "A Multi-Actor Analysis of the QoE Environment", CTTE 7.-1.6.2010, Gent, Belgium;

[5] T. Monath, M. Kind, *et.al.*, "Economical Analysis of Experience-Optimized Service Delivery", CTTE 7.-1.6.2010, Gent, Belgium.

[6] T. Heger, M. Kind, T. Monath, "Drivers, Barriers and Threats for the Integration of QoE Enhancing Technologies within the Access and Aggregation Network", WTC 13.-14.9.2010 Vienna, Austria, (accepted paper);

[7] R. Bradfield, *et al.*, "The origins and evolution of scenario techniques in long range business planning," *Futures*, vol. 37, pp. 795-812, 2005.

[8] D. Mietzner, *et.al*., "Advantages and disadvantages of scenario approaches for strategic foresight," *International Journal of Technology Intelligence and Planning*, vol. 1, pp. 220-239, 2005.

[9] European Commission Information Society and Media, "Broadband access in the EU: Situation at 1 July 2008,"

[10] Gartner, "Next Phase of Growth in Worldwide Consumer Fixed Broadband," 2009.

[11] M. Godet, "Actors' moves and strategies: The MACTOR method : An air transport case study," *Futures*, vol. 23, pp. 605-622, 1991

[12] PricewaterhouseCoopers. (2008, January 31st, 2010). IPTV-Das neue Fernsehen.

[13] IST MUSE, "TONIC tool model A-C with the FTTCab and LEx mixed deployment scenario", 2008.

[14] S. van den Berghe, *et.al.* "QoE driven broad-band access", 2008 NEM Summit, Saint-Malo, France, 2008.

[15] M. de Jonge, *et.al.* "User experience of next generation content delivery: users are developing a new sense for quality", Proc. EURO ITV 2010, June 2010, Tampere, Finland.

[16] R. Huysegems, *et.al*., "Enablers for non-linear video distribution", Bell Labs Technical Journal, Vol. 16, No. 1, June 2011. (accepted paper).